EFFECTS OF STUDENT TRACKING ON ELEMENTARY STUDENTS IN THE
UNITED STATES

by

Claire Kim-Sun Hughes

A Project Submitted to the Faculty of

The Evergreen State College

In Partial Fulfillment of the Requirements for the Degree

Master in Teaching

2013

This Project for the Master in Teaching Degree

by

Claire Kim-Sun Hughes

has been approved for

The Evergreen State College

by

_____

Lester Krupp, Member of the Faculty

June 2013

ACKNOWLEDGEMENTS

ABSTRACT

Negative effects of student tracking on elementary students were examined in this literature review. Student self-concepts were affected by track placement. Teachers perceived lower-track students as being less able to engage in higher critical thinking skills, and relied on subjective measures to group students based on ability. Students made social comparisons to other students within their own track placement. Mixed findings included positive gains in reading in kindergarten when homogeneous ability grouping was used. Further research could include comparative study between groups of students who experienced different lengths of time in tracked classes. Teachers can use flexible, heterogeneous grouping during differentiated instruction.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

**Introduction**

The term *tracking*, refers to the practice of differentiating students based on perceived ability. Desirable/undesirable behaviors, family background, and minority status are all factors that have been used as ways to differentiate. Tracking comes in different forms: between classes, within classes, and between schools. Tracking is different than the term *ability grouping*, although they are often considered to be synonyms. Tracking is tied to curricular differentiation, where low, grade-level, and high distinctions are made. Ability grouping is tied to instructional differentiation, where groups of like peers are placed together so as to increase instructional efficacy. Ability grouping is largely used for reading instruction at the elementary level, while tracking occurs as early as the middle school level and is largely used to determine placement in core subjects like math and science.

The controversy behind the use of tracking is due, in part, to the assertion that within each institution there lies a hidden curriculum which excludes some students from access to an equitable education. "*Hidden curriculum* refers to messages communicated by the organization and operation of schooling apart from the official or public statements of school mission and subject area guidelines" (Cornbleth, para. 1). Hidden curriculum is concerned with the behaviors, the institutionally reinforced values, and the accepted cultural practices associated with schooling. These implicit or unspoken rules of behavior

and social interaction exclude many students and families, with a disproportionate amount of minority and low socioeconomic status students being negatively affected. According to Cornbleth, there are structural and cultural parts of the hidden curriculum. The structural parts include time allotted to different subjects, and the cultural parts include decorations, student cliques, and other social behaviors. Oakes (2005) wrote, "Students can learn *in school* only those things that the school exposes them to" (p. 73).

The education that is available to those who are in the low (below grade level) and middle (grade level) tracks, has come under scrutiny as being less varied, relying more on rote learning, and being staffed by less effective teachers. For example, in math classes, lower-tracked classes encountered rote drill on basic computation skills, while higher-tracked classes encountered more mathematical reasoning and logic skills (Oakes, 2005). This argument demonstrates the issue that is at the heart of the tracking debate: Why are some students learning to think in sophisticated ways while others are still learning through rote memorization? This ties directly into the criticisms that status is conferred onto students in higher tracked classes, and that the academic preparation that they receive is more concerned with higher cognitive functions, whereas the academic preparation that others receive is more concerned with memorization.

Proponents of tracking assert that the use fills the void left by legislation concerned with catching people up, such as the No Child Left Behind act of

2001. Further assertions include that gifted and higher-achieving students are held back in order to level the educational playing field. Homogeneous grouping is believed to be beneficial, where students of like abilities learn and associate with peers. Another term for this type of ability grouping (which results from tracking) is self-contained classroom, and it is polar opposite to a heterogeneous, or mixed-ability classroom. Loss of instructional time is one of the largest criticisms of the use of heterogeneous classes.

Proponents for heterogeneous classrooms assert that the use of student tracking can have detrimental effects on academic achievement and student self-perception. Student tracking is also exclusionary because access to coursework and resources is different among diverse groups of students. Student tracking starts as early as middle school; tracking often originates with standardized testing to determine perceived ability. Then students are placed in tracks, and movement between tracks can be nearly impossible. At the elementary level, teacher perceptions of student motivation, independence, and behaviors, are tied to the groupings that are used in the early elementary classrooms. The practice of sorting students into ability groups shifts quickly into the sorting that could be done later in their educational careers.

Student tracking has longstanding historical roots. Starting as early as 1900, the changing demographics of the nation's population also changed the workforce: increased immigration changed the composition of the workforce. Immigrants often found work in trade or agriculture, because their value as

human capital was largely derived from their country of origin. The goal of early schooling was to Americanize immigrants through deculturalization, or the stripping away of cultural identity for a single, more acceptable, monoculture (Spring, 2011). Society needed a way to differentiate between the education of those who were going to attend college and those who were going to work.

Tracking for high school students increased steadily through the1920s. Intelligence tests made tracking easier and more feasible by yielding quantifiable data as a basis for student comparison. Different students were placed on different tracks. This influenced the peers they associated with, the courses they took, and the preparation they received for their future careers.

Ability grouping was also used as a synonym for tracking, but the rationale for the use of this type of education was to help bring students of similar abilities together; the result, educators thought, would be advantageous for both teachers and students. Educators did not have to differentiate instruction within the same class. Instead, educators would be able to plan for specific lessons catered to the specific needs of their students. The result was a decrease in planning time and an increase in instructional time. For students, ability grouping was considered advantageous because of a decrease in the social labels applied and a potential for an increase in confidence based on being with peers who were all at the same ability level. Students did not have to compare themselves across tracks or within heterogeneous groups, where students with higher abilities may have decreased others' overall self-concepts. Students would be able to compare

themselves with peers within their own classes instead, which could be a less destructive type of comparison. Those who wanted to paint a rosy picture of ability grouping made statements about how ability grouping is good for all levels: high fliers are allowed to soar, low fliers are allowed to rise, and teacher attention is more focused and attuned to student needs. Another argument suggested that, when students are all together in one class, lower students would be neglected in an attempt to move the rest of the class forward, and higher students would be held back in an attempt to keep them with the rest of the group.

There were potential drawbacks to student tracking, which included decreased mobility between different educational tracks; it seemed to some as though the track on which a student was placed could dictate the role that he or she would fill in greater society. Tracking was created out of the intention to align the need for differentiation with the needs of the changing workforce. Not all students were planning to attend secondary education, and people viewed it as being unfair to standardize the curriculum across all students. Those who advocated tracking believed that it was in the best interests of students, especially those who were hesitant to engage in book learning and who might have pursued careers in vocations rather than higher education. The controversy persists because of the potential for negative effects on social and academic achievements. Self-perception, motivation, and self-concept in content area were all potential drawbacks for the use of tracking.

Tracking is not unique to the United States. It is used around the world for

the purpose of differentiating between students and reflecting the needs of the students as well as the workforce. While other countries engage in overt tracking, meaning that the knowledge is known by the general public, the U.S. engages in covert tracking, meaning that the knowledge is unknown by the general public. The U.S. is built upon the principle of personal freedom, and many opponents of student tracking believe that the use of this practice in education violates a student's personal freedom of educational choice.

This paper will address the following question: What are possible negative effects that student tracking could have on elementary age students within the United States? This is a relevant research question because the roots of tracking extend all the way to the early elementary grade levels. An answer is necessary to help settle the debate of whether or not to group students based on perceived academic ability. The heart of the question is who is being left out with the use of student tracking. Since student tracking usually occurs at the secondary school level, it is often during the lower level grades when different indicators, such as behavior or conformity, are first used (along with grades and test scores) to determine groupings. In middle school, students are closely monitored and assessed to help the schools determine which track they will end up on. Academic success is often tied to the effective use of tracking, where higher and lower ability group students are able to socialize and be given differentiated instruction. However, there is often both limited mobility between different tracks and potential social stratification as a result of the use of tracks. Student tracking

is an educational practice that is pervasive and hidden, and it has long-lasting effects for students. It is worth studying because teachers need to know how grouping strategies have a direct effect on the type and depth of education offered to different groups of students.

## Rationale

The possible negative effects of student tracking on academic achievement for elementary students could have a profound impact on the educational future of students involved. Educators are charged with the task of helping equip students with the necessary academic and social skills needed to survive and thrive within the larger society. Elementary education comes under especially close scrutiny because these students are also children. Overwhelmingly, parents, families, and other support group members want to support and protect their children. No one wants to disadvantage a child from the start, so pressure increases for children to associate with the right peers, socialize the right way, and perform academically in the right way. Tracking then becomes an imperative subject of study, because its use could disadvantage some students in order to elevate others.

The use of student tracking, sometimes referred to as ability grouping, constitutes an educational practice that has both pros and cons. Ability grouping refers to instructional differences, including depth, difficulty, and type of subject content, which is differentiated among groups and usually occurs within the same class. When teachers use ability grouping as an educational strategy,

7

effectiveness can include factors like flexibility and duration of grouping in conjunction with the educational practices of the teacher. With a certain amount of flexibility in place, students are able to move between ability groups fairly easily, and within different subject areas. "They [tracking and ability grouping] are also two of the most harshly criticized practices in American education. Both group students of similar achievement levels for instruction, but they differ in how this is accomplished" (Loveless, 1998, p. vii).

Tracking, however, usually refers to the differentiation of students between classes, where they are removed from the heterogeneous class in favor of placement in a homogeneous class with classmates of similar ability levels. Tracking could have a negative effect on social and academic mobility. Tracking is also used more predominantly in science and mathematics courses, where ability grouping occurs across all content areas.

One argument for tracking is that it allows gifted students to be sufficiently challenged. With the rise of legislation like No Child Left Behind (NCLB), the goal had become to help increase learning of underachieving students. An unintended consequence of NCLB was the appearance of leaving behind or neglecting advanced track students. Associate Director of the Gifted Education Resource Institute at Purdue University, Marcia Gentry (2006) wrote, "By conforming and playing the score boosting game, we are denying an entire generation of children quality education while we scramble to raise test scores. NCLB has created an environment in which school administrators have no incentive to concentrate on

educating gifted children or developing talents of any children". Another argument is that teachers do not have to provide as much differentiated instruction (especially when using between-class ability grouping), which decreases overall planning time and could possibly lead to more class time. McCoach, O'Connell, and Levitt (2006) found that the use of ability grouping had the potential for positive gains in reading and early literacy.

One argument against tracking is that mobility is difficult to achieve between tracks. Students placed on lower or middle tracks may have limited opportunity to move upward to the higher track; the requirements from the higher track may be the basis for college admissions decisions and other high-stake future mobility decisions. Students in lower tracks are usually expected to learn more and learn faster, in an attempt to catch them up to other classes. Teacher perceptions of students' ability levels can have a significant effect on the type of instruction, content, and latent messages that students receive. Neumeister, Adams, Pierce, Cassady, and Dixon (2007) discovered that even after participating in professional development training, teachers of potentially gifted students had difficulty recognizing gifted characteristics displayed by diverse students. This strengthens the assertion that perceptions play a major role in track placement.

Another argument against tracking is that it could create stratification between group members identified with different tracks. Social stigma associated with being on any of the tracks could be a legitimate concern for the use of

student tracking. Students who know the rules of the hidden curriculum are less likely to end up in lower tracked class, which is where inequity occurs. Brint, Contreras, and Matthews (2001) studied socialization messages used by teachers and educators during elementary school; the messages were neutral and were presented as directing instead of criticizing or praising. The implicit message was that students were expected to know the rules. This assumption of success tied to implicit knowledge is detrimental to students from backgrounds and cultures that do not rely, or place value on, mentioning rather than explicit telling.

Proponents of tracking argue that current practice differs substantially from practices of the 1920s. Modern tracking, or ability grouping, creates classrooms of the similarly able students and does not necessarily determine career or life tracks. Gifted students are able to interact with others, and those in other groups are able to work together at a level more appropriate for their educational needs. Bain and Bell (2004) found that gifted students rated themselves higher in socially related self-concepts and attributed social success with internal ability more than their high-achieving classmates who were not identified as being gifted.

Opponents of tracking argue that such differentiation is harmful in any form, in that tracking changes the quality of the education a student could receive. This could also lead to different levels of status within the classroom. Social relations and academic achievement are both considered to be negatively

impacted by the use of tracking. The term *detracking* refers to the movement in education in which schools try to eliminate the use of student tracking at an institutional level.

Another rationale for further research on the topic of the use of tracking is that there is insufficient research on the topic available: both qualitative and quantitative. Many peer-reviewed studies were conducted outside of the U.S. which decreases generalizability across countries. Additionally, understanding the potential strengths, weaknesses, controversies, research base, and resources available within the U.S. will help U.S. teachers make more well-informed decisions about instruction within their classrooms and schools.

## Historical Background

Student tracking is a contemporary educational issue with historical roots extending all the way back to the early 1900s. In 1901, Theodore Roosevelt was elected president, and he understood that the regulation of big businesses was necessary. He determined that the abolition of big businesses was not the solution to the nation's problems of the economic inequity and injustice ("American President," para. 4). Industries, such as agriculture, became more prevalent in part because of the changing demographics of the nation's population; immigration was often ebbing and flowing, and each wave brought a subsequent swell in the workforce and in the schools. In 1901, the population of the United States was estimated to be 76,094,000, and by 1920 the population had grown to an estimated 106,461,000 (U.S. Census Bureau, 2000).

11

In April 1906, The Massachusetts Commission on Industrial and Technical Education (MCITE), submitted a report to the Massachusetts Senate and House of Representatives which stated:

They [sons of union members, who were previously recruited by age fourteen] are losing time by beginning so early. You must send them after the elementary schools into these other schools, that they may be initiated into the scientific foundations of the trade they intend to follow, and also from these schools select a trade which fits the peculiar natural ability of a boy (Massachusetts Commission on Industrial and Technical Education, 1906, p.159).

This statement by MCITE implied that elementary and secondary schools were adaptable to include the curriculum needed to create the type of workforce that would have been useful to those associations.

The Smith-Hughes Act (The National Vocational Education Act) of 1917 helped increase the validity of vocational education for those over the age of fourteen who desired to enter a trade such as agriculture. Part of the act was devoted to increasing the requirements for vocational teachers, specifically requiring that they had to have had experience in the vocation in order to teach the subject. The passage of the act was significant because it tied federal funding to fields of vocational study where trade skills were most valuable.

In an attempt to decrease the dropout rates nationwide, the Commission of the Reorganization of Secondary Education (which was appointed by the

National Education Association) published the *Cardinal Principles of Secondary Education*, which stated:

> . . . education in a democracy, both within and without the school, should develop in each individual the knowledge, interests, ideals, habits, and powers whereby he will find his place and use that place to shape both himself and society toward ever nobler ends (Bureau of Education, 1918, p. 9).

This goal supported the idea that there were many functions citizens could serve within society. The authors of the passage implied that, in order increase democratic education for all, there must be other educational avenues for students who were not continuing to higher level education; such coursework in relevant vocational training might have helped them find their places as a contributing members of larger society.

Tracking in the schools began in the 1920s, after high schools made the shift from traditional general preparation for entrance into higher education, to what was termed "career tracking" ("Master Timeline," 2001). Ellwood Patterson Cubberley, former Dean of the Education Department at Stanford University, tried to apply principles of business and scientific management to educational administration. The shift toward a more scientific approach to education, coupled with the rise of intelligence testing, helped change the educational and social landscape.

Tracking was usually involuntary and by 1932, "Seventy-five percent of

large United States cities reported using written standardized intelligence tests to assign students under a tracking system" ("Master Timeline," 2001). Further, the use of tracking placed students on tracks that created differentiated curriculum. Students in each track had limited mobility between tracks.

**Definitions**

Tracking and ability grouping are often incorrectly used as synonyms. This section will outline the difference between the terms tracking and ability grouping. Further, ability grouping can be be broken down to include within class and between class organizational structures.

The term *tracking* refers to the practice of differentiating students based on perceived ability. Main tracks include: high (advanced), middle (at grade level), low (below grade level). These corresponded with the historic labels of "college preparatory, general, vocational" respectively ("Ability grouping," 2002, p.1).

While there are distinctions between tracking and ability grouping, the overlap causes many people to use the terms as synonyms.  Tracking (ability grouping) can refer to either within-class ability grouping or between-class ability grouping. The National Education Association (2010) confirmed that the term *within-class ability grouping* refers to the use of smaller groups within the classroom, when the groups are formed of members of similar abilities. According to the NEA, the term *between-class ability grouping* refers to the more historically practiced idea of tracking, which includes separating students of different abilities into differentiated curriculum.

**Limitations**

The limitations of this paper include the fact that the author examined the use of tracking or ability grouping outside of the United States. Outside of the United States, the use of student tracking was overt. Research results were then generalized for application to United States schools. Research in student tracking within the United States centered on middle and high school grade levels. In order to maintain the scope and relevance of this paper, research discussed here focuses on early elementary and middle school students, although some studies reviewed were conducted with participants who were in high school. These studies were included because of the relevant applications found in the research results which could be generalized to elementary and middle school grade levels.

**Statement of Purpose**

This paper is a critical review of 30 peer-reviewed articles related to student tracking and ability grouping. This paper analyzes methodologies and findings to clarify the issues for practicing teachers and to lay a foundation for future educational research. Student tracking and ability grouping are widely used across many grade levels. Educators increase their effectiveness by being aware of the multiple factors associated with student tracking structures, including student self-concept, teacher perception of perceived ability, potential bias based on socioeconomic status (SES), and social comparisons as a result of track placement.

**Summary**

Tracking students based on perceived intelligence or academic ability has been controversial since 1900. Tracking students is done in a covert way. In contemporary education, tracking could be associated with a lack of transparency in educational practices. Additionally, tracking is often associated with the use of high-stakes testing. In contemporary education, student tracking is a highly charged topic.

During the 1900s, the nation was changing. The population had increased significantly, and there were waves of immigration that flooded the United States with students who had different needs and perceived ability levels. In the early 1900s, there was also a boom in industries including agriculture, and there was the growing recognition that not all students were college bound; therefore they did not all need to be on the same educational path.

The invention of intelligence tests gave school administrators quantifiable data to use for tracking. Students were placed in low (below grade level), middle (at grade level), and high (college bound) tracks. Those in the lowest tracks were able to take vocational courses, which would help them move into their trade as more skilled workers.

The Smith-Hughes Act of 1917 tied federal funding with more skilled vocational teachers, and gave school administrators the incentive to increase their vocational offerings. Anyone over the age of fourteen who was intending to go into a vocation was able to pursue that track.

Tracking also had social aspects. Students within the same track were the main source of social interaction. Students in the same track took the same courses and were presumed to be heading into the same career path. This could have potentially solidified the strengths of these types of relationships. One potential pitfall is that students in one track had very little interaction students in other tracks. Therefore, mobility between tracks was difficult.

Ability grouping is a modified form of student tracking. Ability grouping can occur between-classes or within-classes. One strength of ability grouping is that gifted students are able to be more sufficiently challenged than in a heterogeneous grouping. A second strength of ability grouping is that struggling students are able to receive instruction together, without the additional pressure of trying to catch up with others.

One drawback to this educational practice is the assumption that all struggling students are struggling in the same way. This assumption overlooks the possibility of a diverse group of students. In addition, student associations with different ability groups could be socially detrimental. Many schools are trying to implement detracking measures. One measure is regrouping all students in one heterogeneous classroom and differentiation instruction is used for the group.

CHAPTER 2: CRITICAL REVIEW OF THE LITERATURE

**Introduction**

Student tracking is a complex topic that involves multiple factors. This chapter covers a diversity of topics, which is indicative of the depth of the subject. The first five  studies consider the influence of teacher perceptions. Each research team examined a different relationship between teacher perception and resulting instructional practice.  Types of teacher-student interaction, teacher perceptions of perceived ability, teacher perceptions of gifted or high ability, teacher perceptions of parental socioeconomic status and ability, and teacher perceptions about instructional time spent in preparation for standardized testing, are the main topics addressed in this section.

The section also includes three studies which discuss different instructional methods used based on student grouping. These studies are tied to the assertion that students in different tracks receive differing quality and quantity of educational practices. Interactions within the classroom and student reactions to interest level of subject matter being taught at different track levels are the two main foci of this group of studies.

Student perceptions, which are presented in the next section, refer to intrinsic motivation, self-concept, self-reported ability, and other subjective measures of ability, as reported by students. Six studies are reviewed here. Student perception is a main concern when examining the use of student tracking. Are there less quantifiable, yet equally problematic, results that could

occur from the use of student tracking? Intrinsic motivation and engagement, social comparisons, meta-cognition and test performance, questioning strategies, socialization messages, and student stress and coping mechanisms used during standardized proficiency testing, are all part of this cluster of studies.

Within the section on student perceptions, three studies are reviewed that focus on student comparisons. There are three studies within this grouping. Students often engage in comparisons, but how much is influenced by the use of student tracking? Within-track comparisons (in the same class) and across-track comparisons (between classes), upward social comparison (with classmates perceived as being more academically advanced), and gifted student comparison with high-achieving students, are addressed in this section.

Four studies explore outside factors. This section addresses factors that include parental warmth and communication, gentrification, student mobility and school Adequate Yearly Progress (AYP), and Montessori comparison with traditional education programs. These are all factors that have been tied to the use of student tracking. For example, parental warmth and communication were related to the frequency and occurrence of undesirable school behaviors. Teachers could use conduct problems in their perceptions about student ability, which is a determining factor in later track placement. Gentrification results in oversaturation of some socioeconomic groups of students. Students may receive varied quality of instruction, which could either advantage or disadvantage them. While suburban schools are receiving the more affluent families' students,

neighborhood schools are struggling for resources, including quality educators. Student mobility refers to actual student movement between different schools. The decision-making process of changing schools could be directly linked to increased social mobility or perceived higher academic curriculum opportunities. The result is that student mobility can have a large impact on AYP, a yardstick against which schools are measured and a decisive factor for federal funding. These studies highlight important changes that stem from and have an effect on both tracked and detracked schools.

Two studies found that positive results could be attributed to the use of ability grouping and tracking. Studies focusing on both improvement in early reading and literacy, and on positive student self-perception about ability grouping are addressed in this section.

Two other studies concluded that negative results could be associated with the use of tracking or ability grouping. These studies explore delineations around entire groups of students, different instructional methods, and the hidden nature of the use of tracking in the United States, as compared to other countries. Additionally, one study found both positive and negative findings about the use of ability grouping and tracking. Instructional methods, learning environments, and amounts of material taught vastly differed and yielded different results among grouped and non-grouped classes.

The final section is composed of studies about detracking and alternatives to tracking structures. Four studies comprise this section. Alternative systems,

such as Programmed Learning Sequence (PLS), cooperative learning, reduction of prerequisites to allow movement between tracks, and approaches which include legal action and institutional change, are all included. These studies take up issues of how schools can provide for the range of student abilities and interests without resorting to tracking. How do schools detrack? With so many schools using tracking, what does it look like when tracking is voluntarily decreased or eliminated? How could that work? These are the crucial questions that will influence administrators and educators alike, and will ultimately inform classroom practices.

## Teacher Perceptions

This section discusses the powerful effect that teacher perceptions have on classroom interaction, group placement, and instructional practices used among diverse groups of learners. Variations included praise/criticism, types of questioning, and teacher responses about perceived student ability. This has direct implications on future teaching practices because reflective practitioners need to be able to see the full picture of the classroom culture, both explicitly stated and implicitly understood.

Teacher perceptions influence decisions for instructional grouping. Haller and Davis (1981) investigated teacher perceptions based on parental SES, they found that parental SES was not the most influential factor in teacher decision making. Haller and Davis discovered that teachers relied most on perceived student ability when recommending student placement in one of three different

21

reading groups. Haller and Davis' study brought awareness to the variable of teacher perception in the classroom.

In addition, Parsons, Kaczala, and Meece (1982) studied the quality of student-teacher interactions and how interactions affected student self-perception or self-concept of ability. Upon studying math classes, researchers found that teachers gave female students less overall criticism than male students. Parsons et al., also found that teacher criticism had the greatest impact on student self-perception of ability. Parsons et al., drew attention to the importance of studying socialization messages students received in the classroom from their teachers.

Similarly, Leder (1987) used a case study methodology to investigate the quality and quantity of interaction between teacher and students of perceived differing abilities. Leder determined that teacher perceived higher achieving students were more frequently addressed by the teacher. Additionally, teacher perceived lower achieving students were afforded less wait time when asked questions.

Neumeister et al. (2007) examined teacher perceptions of gifted students in classrooms with high populations of minority and lower SES students. Neumeister et al. used a qualitative survey to expand the scope of study of teacher perceptions of ability to include giftedness.

Teacher perceptions influence instructional practice. The following two studies investigated how different instructional methods were used based on

22

student groupings. These three studies tie into the argument that students receive different instructional methods across different tracks. Eder (1981) examined the variety of teacher-student interaction across different levels of ability grouped students. While Eder used micro-analysis of a single classroom, the researcher's longitudinal study found significant differences in teacher-student interaction. The teacher corrected students in the lower reading group twice as often as those in higher groups, and employed behavior management strategies twice as often as those in higher groups.

Kersten (2007) explored the connection between community valued literacy skills and school valued literacy skills. Even with the small sample size of seven participants, Kersten's conclusion that students who were considered higher achieving were assigned higher cognitive assignments than their peers was startling.

External pressures affect instructional practices. The following two studies are related to teacher perceptions of external expectations and how these perceptions influenced their   practices. One criticism of lower track classes is that teachers only teach the materials that might appear on standardized tests. Monsaas and Engelhard (1994) found that when teachers reported higher levels of pressure to prepare students for standardized tests, teachers would adjust instructional practice to include more test-taking strategies.

Researchers Waxman, Huang, Anderson, and Weinstein (1997) statistically analyzed how teachers used different instructional practices between

higher performing and lower performing urban elementary schools. Similar to the findings of Monsaas and Engelhard, researchers found determined when additional pressure was placed on teachers, instructional practices reflected those changes. In lower performing schools, students spent more time in whole-group instruction, spent less time interacting with each other and the teacher, and spent less time in independent study than did students in higher performing schools.

Haller and Davis (1981) focused on the question of how strongly factors including parental socioeconomic status (SES) were relevant to teachers when making reading grouping decisions, and performed quantitative analysis on the qualitative data. Researchers collected data from in-person, informal teacher interviews. Researchers tried to determine if a correlation or relationship existed between students' attributes and teacher recommended group placement. Haller and Davis  provided teachers with information about parental socioeconomic status (SES) and individual student test results. Regression analysis was used to control for SES and zero-correlation.

Researchers chose reading because that subject utilized the most ability grouping at the elementary grades. Haller and Davis (1981) discussed Rist's argument that teachers had a tendency to group students based on perceived ability, which is measured against the teacher's personal belief about the "ideal student" (p.172). The teacher would make comparisons based on how close or how far way the student was perceived to be relative to the ideal. Haller and

Davis (1981) also noted that Rist drew attention to the fact that the highest level of teacher effort and attention was devoted to students placed in the highest reading groups.

Haller and Davis (1981) gathered data from 37 fourth-, fifth-, and sixth-grade classrooms from five schools in central New York state. Their study involved 960 students, and the sample consisted of wide SES distribution. Researchers noted that external validity was limited due to a lack of racial variation within the sample size; less than 1% of students in the studied district were African American. Therefore, study results may not be generalized across more diverse populations with confidence.

For comparison, student data was gathered from the Iowa Test of Basic Skills. Students used a modified version of Gough's Home Index, which consisted of 22 items, including topics of family possessions and parental educational attainment; this index was used to gather parental SES data.

Thirty-seven teachers were given note cards with student names written on them. Teachers were asked to engage in a card-sorting activity, where each name needed to be sorted into one of three reading groups. Haller and Davis (1981) asked teachers open-ended, informal, interview questions during the card-sorting activity, and tried to have teachers elaborate about student placement in different groups. Researchers used neutral questions to focus on the reasons behind the visible student placement.

Teacher comments were coded according to similar meaning groups, and

researchers narrowed the scope to five major groups. Two independent coders had a 91% inter-rater agreement among the five groups (Haller & Davis, 1981). Responses were coded based on *+, -, 0* symbols, where a plus-sign meant a positive tone was used, a minus-sign meant a negative tone was used, and zero meant a neutral tone was used. Teachers commented about the factors of reading ability, general ability, and work habits most often. Twenty percent of the total comments fell into one of the three categories, and were applied to 75% of the students. In 25% of their comments (affecting 240 students), teachers did not mention reading ability when they discussed grouping. During the card-sorting activity, teachers were not asked questions about the original reasons for grouping at the beginning of the year. Researchers also collected data later in the school year, when knowledge and relationships were stronger than they were at the beginning of the year.

Haller and Davis (1981) found that when teacher judgment was used to form the groups, the degree of SES segregation increased about 20%. Strength of teacher perceptions was calculated by subtracting negative remarks from positive remarks, and was used to determine the influence of perceptions about parental SES and reading skills. Haller and Davis (1981) reported the factor of general ability and achievement had a .58 correlation with teacher-recommended group placement ($p < .01$). The factor of reading ability and achievement had a .51 correlation with teacher recommended group placement ($p < .01$). Teacher recommendation for placement was strongly correlated with perceived ability

(almost twice as strong as other variables). Teacher recommendations for placement and SES were weakly correlated (.34), and test scores were not strongly correlated with perceived reading ability (.38). The weak correlation of approximately .3 indicates that the variables were not readily related, meaning a change in one did not necessarily relate to a change in another.

The results of this study show a  weak correlation between parental SES and teacher recommended reading group placement (.11), which suggests the existence of little bias based on SES (Haller & Davis, 1981). The study had reliability, and the research methodology was explicitly stated so that it could be predictable from one study to another. The study also had objectivity and transparency so that outside parties would have been able to observe the research. One potential threat to reliability and external validity (noted by researchers) was racial homogeneity: Caucasian students were overrepresented. Therefore, the results might not be able to be generalized to other settings with more diverse student populations.

Parsons et al. (1982) were concerned about the quality of student-teacher interactions, and if variations in teacher-student interaction patterns had an effect on student self-concept about ability, where student self-concept was tied to student expectation for future academic performance. Researchers also examined how socialization messages were communicated within the classroom. This qualitative study involved 427 students selected from 17 different math classes from fifth, sixth, seventh, and ninth grades. Parsons et al. noted that they

coded teacher-student interactions, and four-trained coders coded 10 sessions per class. Researchers collected and coded observational data during two months in spring 1978.

Students were asked to fill out a questionnaire (unnamed scale) based on a 1 to 7 Likert-type scale (Cronbach's alpha referenced but not given) that asked about self-concept of ability, student concept of the task, and student expectancies. Parsons et al. (1982) also administered a teacher questionnaire which comprised six questions about the teacher's students; teacher expectancies were calculated based on the mean from two different questions: 1) How well will _____ do in advanced math? and 2) How good is _____ at math?

After researcher coding, the results indicated that nine different variables were related to gender. Females students had lower future expectancies ($F$ = 5.555, p. 019), thought that math was harder ($F$ = 4.10, $p$ = .044), had less total criticism ($F$ = 8.16, $p$ = .005), had less work criticism ($F$ = 6.56, $p$ = .011), had less conduct criticism ($F$ = 6.21, $p$ = .013) and thought that less of their response opportunities had been modified ($F$ = 7.00, $p$ = .009), than did their male classmates. Additionally, female students asked more questions ($F$ = 10.84, $p$ = .001), especially questions related to procedure ($F$ = 8.74, $p$ = .003) than did their male classmates.

Parsons et al. (1982) found that teacher-student criticism had the strongest effect with regard to student attitude. Teacher-perceived high and low

male students received the biggest differences in teacher-student interaction. Also, female students reported a higher impact of the changing nature of criticism and praise than did their male student classmates.

Parsons et al. (1982) included the history of the topic of socialization attitudes in the classroom, and recognized that little previous research existed. Researchers noted that one factor that could have an influence on future studies is frequency of occurrence of criticism and praise. Criticism and praise frequencies were low in their study, so future studies with different frequencies could yield different results. The study had greater reliability based on a higher number of coders and level of inter-rater reliability, with a range between 75% and 86%. High percentages of inter-rater reliability help increase confidence in the collection and analysis of qualitative data.

Student-teacher interaction was the primary focus of Leder's (1987) case study. The researcher examined three different aspects of student-teacher interaction: 1) interaction by gender, 2) interaction across subject areas (math, language, science), and 3) interaction with teacher-identified low- and high-achieving students. Leder focused on 30 students from a sixth-grade classroom in Melbourne, Australia. Seventeen participants were girls, 13 were boys, and the teacher was male. Leder (1987) focused on the top two students and lowest two students. The researcher used both test scores and teacher designation to determine student achievement status. The top two students received more teacher interaction and were asked questions more frequently than were the

lowest two students. When the teacher called on lower-achieving students during class, after waiting for a short period of time or encountering a string of incorrect answers, the teacher would then call on higher-performing students to get the correct answer. The teacher asked product-related questions four times more often than process-related questions. Leder noted that there were no differences found in the length of interaction across achievement groups. Leder's study demonstrated how teacher perception of student ability changed the quality and quantity of interactions with students. Proponents for detracking measures argue that students in lower tracks receive education that differs in quality than students in higher tracks.

The small sample size, the lack of demographic information, and the small number of total observations decreased the credibility of the study. The study could be transferable, in that observations could be carried out in contextually similar settings and possibly yield the same findings. The researcher did not mention if findings were consistent with other studies. Leder (1987) did not include information about how data was collected and evaluated, which decreased study's confirmability.

Neumeister et al. (2007) examined three foci in their research. Researchers focused on teacher perceptions from teachers of diverse classrooms with high populations of minority and lower-SES students. Researchers also focused on unique characteristics shared by underrepresented groups in gifted services; Neumeister et al. were interested in determining the

level of support students received in their placement within gifted programs. Teachers were heavily involved in the student referral process, and questions were developed to help assess the types of characteristics or qualities teachers associated with potentially gifted students. Researchers were curious as to whether or not teachers of minority and economically disadvantaged students held the same perceptions when it came to identifying gifted students as they did with identifying students from the social majority.

The study was conducted during 2003, and 27 teachers participated in the study. Selection criteria included teachers being involved in Project CLUE (Clustering Learners Unlocks Equity, associated with Ball State University) and teaching in a gifted cluster. Twenty-one of the 27 teachers (about 77%) had received additional information about instruction and gifted students during a 4-day summer institute.

Teachers were contacted by telephone and e-mail to encourage participation by filling out surveys. Of the 27 teachers who responded, 26 were female and 25 were Caucasian. Twenty-three teachers self-reported about teaching experience, where the mean was 17.4 years of teaching and the range was from 3 to 32 years. Teachers had a mean of 4 years of experience teaching gifted students, and there was a range from 1 to 15 years of experience. Teachers reported that during 2001-2002, 184 students were identified for gifted services, and this had been determined when the students had completed second grade; over half (58%) were African American.

Neumeister et al. (2007) coded survey responses, developed coding categories for different types of responses, and recorded the frequency of occurrence of each type of response. The first survey question asked teachers to provide personal definitions about giftedness and characteristics they associated with identifying giftedness. The second question asked teachers to identify any concerns they had about student identification as gifted and receipt of gifted services. A third question asked teachers to list the characteristics of their personally identified gifted students.

For the first question, the highest frequency of answers was related to self-motivation or independent learning as being a strong indicator of giftedness. Twenty-one out of 27 responses (about 78%) listed being self-motivated or an independent learner as the top characteristic. The second highest ranking characteristic was the student demonstrating faster learning or easier learning in different content areas. Neumeister et al. (2007) were concerned that teachers may have been equating productivity with giftedness (Neumeister et al., 2007, p. 492). Only four of 27 teachers (about 15%) reported that the characteristics of being artistic, loving to read, being bored, or showing disinterest may be characteristics of giftedness. The researchers inferred that teachers may have been narrowly focused on characteristics that downplay cultures in which other, less readily accepted, gifted characteristics may be present.

In response to question two, teachers expressed concern about making a distinction between high-achieving students and actually gifted students. The top-

ranked concern in the ability category was that a student had deficient skills in one content area, which could be reading, math, or communication; the frequency of this response was 16 of 27 (about 59%). The second highest concern about ability was that a student grasped concepts in a slower way than other identified gifted peers, with 9 responses (about 33%). For work habits, teachers were concerned that students were not motivated (frequency of 13), or were underachieving (frequency of 10).

Teachers expressed concerns about 63 out of 184 identified gifted students (about 34%). Researchers created Tables 2 and 3 (p. 488 and p. 490) which displayed the characteristics the 27 studied teachers commented about when expressing concerns about gifted students. Frequencies are based on the number of teachers who commented in each of the categories, referred to as "Concerns" (Table 2, p. 488) and "Characteristics" (Table 3, p. 490). Thirteen teachers commented about students who could qualify for gifted services based on the student work habit of *not motivated.* Sixteen teachers noted *student perceived deficiency in one content area* as another concern for students who qualified for gifted services (Table 2, p. 488).

Eleven teachers listed *higher-level reasoning and problem-solving* as ability based characteristics of the students they were concerned about who had been identified as gifted. Six teachers answered that *skills are lower than other identified peers* was the basis for their concern of students who had received gifted services.

33

Neumeister el al. (2007) concluded that teachers of minority or economically disadvantaged students had a narrow focus about characteristics of giftedness. The researchers suggested that extensive professional development could be useful, so that teachers could develop a more multicultural understanding about what giftedness looks like and sounds like across different groups (Neumeister et al., 2007, p.493). Researchers also expressed concerns that behavior problems or perceived underachievement were contributing to the narrow focus used by teachers.

The study has transferability, although all teachers who participated in the study were pursuing graduate-level education. Therefore, it may be challenging to predict findings if the demographics of sample population changed. The study was credible because the research design helped increase confidence in the ability to produce findings that were believable and convincing. Researchers did not include information about inter-rater reliability when coding and categorizing responses, which decreased the study's confirmability.

Eder (1981) used micro-analysis to investigate how teacher-student interaction varied between different levels of ability grouping during reading instruction, and conducted research through participant observation during a longitudinal case study of an individual teacher, a single first-grade class, and individual students. Eder used participant observation, field notes, and videotapes to record the results of the one-year longitudinal case study; the researcher decided on the length of one year because she hypothesized that the

length of time would help identify changes in teacher or student behavior. Eder focused on changes in the teacher-student interactions during group lessons. The researcher chose a first-grade class because it was the first experience with school and the first experience with formal grouping practices.

Eder (1981) used observations three days each week for the entire academic year. Observation periods lasted about three hours and included both morning and afternoon teaching sessions. Reading was selected because up to two hours of the school day was focused on formal reading instruction, while only about five to ten minutes was dedicated to formal math instruction. Thirty-two reading group lessons were taped and coded, and the researcher used a sociolinguistic approach to analyze the data. Coding categories were determined by subjective measures such as observed non-verbal cues and verbal interactions. Eder and a second, independent coder had inter-rater reliability in a range from 84% to 95%.

Before the school year began, the classroom teacher consulted with kindergarten teachers to predict how students would do in reading and to solicit perceptions about where students could start. During the first week of instruction, the first-grade teacher assigned students to four groups; the duration of placement in groups was not mentioned

Eder (1981) discovered that low reading-group members spent 40% of the time not paying attention during reading turns, while higher-group members spent 22% of the time not paying attention. Lower reading group members spent

a greater amount of time engaging with inattentive behaviors (t = 6.23, p < .001). The teacher used more than twice as many management strategies for the two lower groups than for the higher groups, with 157 instances and 61 instances, respectively (t = 4.12, p < .001).

In addition, lower groups were corrected or disrupted during reading turns twice as often as those in higher groups (30 times and 16 times, respectively). The teacher asked perceived nonreader students questions 41 times in the lower groups, while only 21 times in the higher groups. In lower groups, reading violations occurred 153 times, and only 78 times in higher groups. Teacher perception about student abilities in higher groups was much more favorable than of student abilities in lower groups. Eder (1981) speculated that labeling theory had been used, and those students who had been placed in low groups had a higher tendency to act in the way that aligned with the label.

The study had some dependability, where the findings were consistent with previous studies. The study also had some credibility, where the methodology and conduct increased confidence that the study would yield believable and convincing findings. A threat to credibility, transferability, and confirmability was that there was no information provided about the school, the students, or the teacher, which are relevant factors if further research is going to occur. Eder's (1981) study had reliability strength in that the computation used for the t-values was provided, which included definitions for each variable used.

Are forms of community literacy aligned with in-school literacy? Kersten

(2007) was interested in whether the community literacy of urban elementary school students was valued within the school setting. Kersten conducted a case study through in-person, semi-structured interviews of seven female fifth-grade students. Kersten had worked with the classroom teacher before, so students were familiar with researcher's presence in the classroom.

The school was an urban elementary (K-5) school. Teachers identified and selected participants according to their willingness to participate, and the willingness of parents to give consent for student participation. The original group selected was of 12 students, but it decreased to seven students who had consent forms signed by parents. Sample size of seven students was ethnically similar to and representative of the overall school population. Two students were European Americans, three were African Americans, and two were self-identified as students of mixed European and African American descent.

The seven students formed the Literacy Social Action Group and were given assignments that included formally interviewing community members about literacy, keeping records of different tasks that required literacy skills, and bringing in photos along with other examples of literacy in the local community.

Kersten (2007) established the purpose of the literacy group during the first meeting: to explore reading and writing activities in school, at home, and in the community. Students demonstrated a clear delineation of social cliques at the beginning, but over time, students became more interdependent on each other. Students were asked to conduct formal interviews with community members, and

they were given the role of "ethnographic researchers."

Kersten (2007) asked students to record all reading and writing activities that they participated in; one activity included listing places that students frequented more than once a week and then documenting evidence of reading or writing present (e.g., grocery lists for the store, receipts from fast food restaurants). Religion played a large role of literacy in the community. Church and church activities were noted by many of the participants; five out of seven of the students had some sort of record of religious or religion-associated activity in their journals.

Students mentioned food shopping, eating out, journal writing and creative writing as encounters with literacy. Students also mentioned reading and writing habits of family members; most students engaged in similar types of reading and writing habits. Five of seven participants talked about reading with younger siblings outside of school. One student mentioned playing school, where tasks included taking attendance and writing down numbers and letters. Most participants, "mimicked parental literacy practices, particularly of their mothers" (Kersten, 2007, p. 146).

While different social factors contributed to literacy outside of school, Kersten (2007) discovered stark differences within the school. One student was part of the advanced (gifted) program and attended advanced classes on four days each week. Her answers about the type of reading and literacy she engaged in were very different from those of her classmates. She discussed

specific, interesting texts that she read and what she understood given what she had read, in contrast with the other students who listed their favorite reading time as SSR, because it was choice reading. When asked about what they were learning in class, six students wrote generic responses (e.g., science, math, textbooks), while their advanced classmate wrote about Archimedes and his contribution to science. Additionally, another student was part of the gifted and talented program, but who was not one of the study participants. Both students who were part of the gifted and talented program had parents who both held college degrees. The researcher inferred that this was due to a tracking system based on social reproduction.

Kersten (2007) also found mixed results from the semi-structured interviews. Some students said that literacy was the same between home and school, and others said that they differed. One student did not see any connection between school and home literacy. Kersten concluded that students would have been more engaged if they had been part of the decision-making process, and if they had been asked to share about their own literacy experiences outside of the classroom. Kersten's study had low transferability, due to the very small sample size and isolation of sample to one particular school. Also, since it was a sample of convenience, this would need to be taken into account for similar studies; researchers may struggle with recreating a similar demographic and SES with a larger sample size and in a different geographic setting. The study had credibility because of Kersten's (2007)

discussion about her data collection and research methodology.

Monsaas and Engelhard (1994) researched teacher attitudes toward standardized testing practices. Researchers collected qualitative data through voluntary participation questionnaires, and examined the data for correlations between different variables. The participants were 186 classroom teachers who were pursuing graduate work in Georgia; all were from the metropolitan Atlanta area, and the participant sample comprised both small-town and rural districts from the northwest and western areas of the state. School SES was calculated by the parameters of the state's free and reduced lunch program. Of the participants, 96 teachers were from the elementary level, 57 teachers were from the middle and junior high level, and 33 were from the high school level. Participants included 158 female teachers (85%) and 28 male teachers (15%).

Researchers used questionnaires to gather teacher data. The Testing Practices Instrument (TPI) was used to explore three different variables: (a) teachers' perceptions of exerted pressure on them to increase classroom test scores (labeled as variable: pressure), (b) teachers' attitudes about current testing practices (labeled as variable: attitude), and (c) teachers' use or lack of use of testing practices (labeled as variable: behavior).  Attitude and behavior scales had 16 items, and the behavior scale was modeled on the one that had been used by Mehrens and Kaminski in 1989 (as cited in Monsaas & Engelhard, 1994, p. 471). Researchers used a 4-point scale, from *definitely* to *not at all*, to measure the attitude variable. Unfortunately, researchers did not indicate which

40

end score was related to frequency of use. Researchers used a similar 4-point scale to measure the behavior variable, with a range from *frequently* to *not at all*, and no indication about which endpoint corresponded with the answer. The pressure variable was measured using a scale comprising three questions which inquired about parents, school administrators, and district administrators in relation to raising test scores. The pressure scale also used a 4-point scale from *none at all* to *a great deal*, with no specific designation about corresponding endpoints and values. For the pressure scale K-R 20 = .74 (reliability), for the attitude scale K-R 20 = .72, and for the behavior scale K-R 20 = .78. Monsaas and Engelhard noted that they had first added together all the response items related to each scale, in order to arrive at scores for each scale.

Researchers determined that there was a negative relationship between attitude and behavior scales, $r(184) = -.337, p < .001$. Researchers inferred that teachers did not readily report behavior that could be considered to be cheating. Researchers made the assertion that elementary school teachers were more likely to engage in these types of behaviors. However, caution should be used because researchers provided no data to support their hypothesis about elementary school teacher practices. Factors of the grade level taught and behavior also had a negative correlation, $r(184) = .195, p < .01$. Monsaas and Engelhard (1994) applied a linear model of statistical analysis to examine effects and interactions. Pressure to increase test scores has a significant effect on testing practices behavior, where SES, grade level, and pressure were controlled

for, F(1,158) = 7.06, p < .01.

Elementary teachers engaged in more test preparation behaviors than did teachers at other grade levels. When teachers felt that pressure to increase standardized test scores increased, there was a corresponding positive correlation with behavior, or the greater the likelihood that teachers engaged in more test prep activities. Data also suggested that teachers in lower SES schools tended to engage in more test prep activities than their colleagues in higher SES schools did (Monsaas & Engelhard, 1994). Researchers noted, "...Elementary students typically have one teacher for all their classes...the teacher might feel more personable responsibility for low test scores, and therefore engage in more practices designed to increase students' test scores" (p. 472). A weakness of this study was that it was a sample of convenience, which included teachers all at the same educational attainment level. Since all teachers were pursuing graduate-level study, this limits the generalizability to populations of teachers who were not pursuing graduate level study. The researchers also excluded the factors of gender and race from the study because of the disproportionate ratio of women to men (158:28) and the disproportionate ratio of Caucasian teachers to African American teachers (167:19). Monsaas and Engelhard (1994) wrote, "Furthermore, there was no rationale for hypothesizing differences based on these characteristics" (p. 470).

In a quantitative study, Waxman, Huang, Anderson, and Weinstein (1997) addressed the research question: How do classroom processes differ between

Effective/Efficient (E/E) and Ineffective/Inefficient (I/I) urban elementary schools?

Researchers used self-report surveys and performed statistical analysis on the data. Researchers had a participant pool to choose from, which consisted of 167 urban elementary schools in a district in the south-central region of the United States. A sample size of eight schools was selected, where four schools were randomly selected from the top quartile (labeled E/E schools) and four schools were randomly selected from the bottom quartile (I/I). Schools were identified using Data Envelopment Analysis (DEA), which was originally used by Anderson and Walberg in 1994 (as cited in Waxman, et al., 1997, p. 49), which measures efficiency. DEA uses a scale of 0.0 to 1.0, where 1.0 equals full efficiency. DEA was used to estimate how efficiently schools were using scarce resources to produce desirable outputs, like standardized test scores. For example, if an I/I school was had an efficiency measure of 0.30, then the school was only 30% as effective as other schools that have the same demographics and were considered to be E/E schools. If a school was operating at less than a 1.0, it was labeled an I/I school by researchers.  Approximately 15 classrooms were chosen from each school, and researchers focused on reading or math. Six students were randomly selected from each class for observations. Researchers gathered a total of 459 observations from grades three through five.

Researchers used multiple regression analysis to increase the reliability of the use of DEA. Achievement test scores and initially reported demographics were additionally used to help further confirm I/I and E/E school status.

For E/E schools, 72% of instructional time was spent on whole-class instruction, 20% of instructional time relied on independent study, and 8% of instructional time relied on small-group instruction. Researchers noted that students worked independently 21% of the time, interacted solely with the teacher 70% of the time, and interacted with other students 8% of the time. Additionally, 98% of tasks given to students were assigned by the teacher. Researchers noted that for 90% of the observation time, students were on task.

For I/I schools, 81% of instructional time was spent on whole-class instruction, 14% relied on independent study and individual work time, and 5% on small-group instruction. As with the E/E classrooms, virtually all (99%) of tasks given to students were assigned by the teacher. Students did not interact, but instead were doing individual work, for 42% of the observation time. Students interacted with the teacher 47% of the time and with other students 11% of the time. Students spent 52% of the time watching or listening, 17% of the time working independently on written assignments, and 6% interacting with others. Waxman et al. (1997) noted that 93% of the observation time, students were observed being on task.

Waxman et al. found that main differences between the E/E and I/I schools occurred in the classrooms. Students in E/E schools interacted with their teachers 70% of the observation time, where students in I/I schools interacted with their teachers 47% of the observation time. In addition, students in E/E schools spent more time working in small groups or individually. Students in I/I

schools spent more time working together in one larger group. An inference is that students in I/I schools spent more time in classrooms where direct instruction was the primary instructional method used. Direct instruction usually involves a lecture style introduction of material, as opposed to student-discovery types of models. Researchers wrote, "Active student learning is another critical instructional process that improves student outcomes, yet students in the I/I schools were observed merely watching or listening in the classroom more than half the time" (p. 56). While the differences in percentages between E/E and I/I schools appear to be slight, students from E/E schools reported more supportive learning environments and more aspirations than did students from I/I schools. Research findings indicate that instructional methods have an impact on student self-concept and academic performance. Within the context of the larger subject of student tracking, I/I schools and E/E schools could be considered tracked schools, where the I/I schools represent the lower tracked classes and the E/E schools represent the higher tracked classes. Instructional methods varied between the tracked schools in the same manner in which instructional methods varied between different ability groups in other research studies.

Multivariate Analysis of Variance (MANOVA) was used to determine the impact of an E/E or I/I on school factors such as interaction, setting, and types of activities. Student self-surveys revealed differences on scales related to social self-concept, academic self-concept, teacher support, and student aspirations. Students from I/I schools rated themselves higher in social self-concept scales

than did students in E/E schools. Students from E/E schools rated themselves higher on scales related to academic self-concept, and student aspirations. Waxman et al. (1997) reported five of the variables they examined were not statistically significant: achievement motivation, involvement, task orientation, rule clarity, and satisfaction. Researchers commented that one limitation could come from the fact that most teachers in the observed classrooms were experienced, had advanced degrees, and were skilled at keeping students on task.

## Student Perceptions

Student perceptions about factors from intrinsic motivation to self-concept in subject areas can be influenced by the type of group the student associated with or was placed in. This is directly tied to teachers' choices about teaching practices because teaching a lesson entails much more than just lesson planning; it also includes evaluating and anticipating who may be excluded from a lesson.

Students are aware of how they are perceived by peers and teachers, and are also aware of how they are meeting overall expectations. Brint et al (2001) studied different socialization messages conveyed to students in the classroom. Researchers reported that students were left to make inferences about classroom expectations during a majority of the observational time, where teachers would present expectations in a neutral way rather than as an explicit statement. This study fits in with the argument that students can have limited

46

mobility between tracks due to the hidden curriculum present in some institutions. Skybo and Buck (2007) discovered that some students had physical symptoms in response to stressors, including proficiency tests. Proficiency tests can be used as a factor in deciding student track placement. One implication of the research findings is that students are aware of that the pressure placed on them to have higher proficiency test scores, and the importance placed on performance.

How do students associate themselves with track placement? Butler (2008) analyzed quantitative data to determine if the use of student tracking affected student self-perception of ability and student motivation to seek help. Surprisingly, students in higher tracked classes avoided seeking outside help more than students in lower tracked classes, which will be covered in more detail in this section. In addition Hulan (2010), found that students were often inhibited to ask clarifying questions or engage in deeper examination of the texts they were reading when the teacher was leading the group discussion. Interestingly, when the teacher was absent, students usually remained on task and discussed that reasonableness of answers. This inhibition could have resulted from the understanding that there are so-called right and wrong answers.

Students can be metacognitive about what they are doing while taking tests. Krebs and Roebers (2010) discovered that eleven and twelve year old students were able to determine how difficult a test question was and how accurate an answer they had given was. This is related to Butler's findings of high student awareness of the implications of asking for help. In Butler's study,

students understood that asking for help could change the way they were perceived by others. In Krebs and Roebers' study, students understood that allowing incorrect answers to remain could change the way they were perceived by others. Crumpton and Gregory (2011) studied student intrinsic motivation. Crumpton and Gregory found that when students felt academic subjects were relevant to their lives, they self-reported higher levels of intrinsic motivation. Research findings indicate that creating a relevant and engaging framework when introducing subject matter could help motivate students to perform at a higher academic level.

Researchers Brint et al. (2001) used methods of classroom observation, teacher and principal interviews, and researcher review of curricular materials, to uncover how socialization messages are conveyed and enforced in primary schools. Researchers gathered data from second-grade and fifth-grade classes, and used these two different grade levels to examine the effects of changes in socialization between grade levels. Researchers examined 64 classrooms in southern California in 1998, with follow-up surveys in 1999 and 2000.

Brint et al. (2001) observed and coded socialization messages that teachers sent to students during a daily one hour whole-class instruction period. One potential drawback is that researchers stated they had a "high level of confidence in the rank order" (p. 160), but did not state the level of inter-rater reliability. Teachers and principals were interviewed after the observations, for about an hour. The 64 teachers used in the study affirmed that they had not

discussed the study with anyone, including colleagues. The study also used curriculum materials to determine the value-relevant messages contained in the classroom texts.

Researchers reported 1,179 total references, with an average of 18.4 references per observation period. Of the 1,179 references, 986 (84%) were related to the basic organizational controls of orderliness (839 references) and industriousness or hard work (147 references). Teachers sent messages about modern values like individual uniqueness 0.4% of the time (four references), about respect for group differences 0.3% of the time (three references), and about respect for own group culture 0.2% of the time (two references). Cooperation was mentioned 2% of the time (21 references), while honesty was mentioned 0.6% (seven references) and courage was mentioned 0.9% (10 references). The study results shed insight into the types of message conveyed to students about the standards governing behavior and values within the school setting.

Socialization messages were also delivered in a neutral way, instead of through direct criticism or praise. Teachers indirectly conveyed behavioral expectations 56% of the time, and praised or criticized 18% and 15% respectively. Brint et al. (2001) examined the existence of a hidden curriculum in routine classroom practices. Students had to rely on making inferences to understand invisible norms.

The study had credibility, in that researchers described data-gathering

procedures and how coding categories were analyzed. Brint et al. (2001) noted the fact that raters had coded some items differently than others. For transferability to other contextually similar situations, the geographic region of southern California was chosen for a specific reason. Namely, researchers noted, "[Southern California] is important, since some practices that are characteristic of the schools we studied, such as the use of token economies to encourage conformity, may not be as prevalent in other states or regions" (p. 159). Therefore, the research method would have to be changed so as to account for different institutional factors. The study had limited dependability because few previous studies existed about how socialization norms are presented in primary school classrooms for comparison.

Skybo and Buck (2007) were interested in how stress and coping responses emerged in school-aged children around times of proficiency testing. Since proficiency testing is often relied upon for track placement, the study is relevant to the research question about the possible negative effects student tracking has on elementary-age students. Researchers conducted a longitudinal, repeated-measures qualitative study. Researchers also used in-person interviews, questionnaires, and case study observation to gather data. Data was gathered during the 2003-4 school year, and only fourth-grade students were used to help control for variables in data. The four data collection points were in October (labeled T1), February (T2), March (T3), and a posttest that occurred 1 month later, in April (T4). Twenty-seven students were part of the study sample

to ensure that repeated measures were reliable and to account for moderate changes.

Researchers began their data-collection by having family members fill out a demographic form about the background of students and any significant changes that may have occurred in the last six months. Then, four different instruments were used to collect data. Researchers used all four instruments at each of the four data-collection points. The first instrument, Proficiency Test Stress Visual Analog Scale (VAS), was conducted when students were engaged in discussions about stress; students were asked to rate stress on a scale of 0-10 (10 was most stressful) about fourth-grade proficiency examples. Researchers used the open ended question, "What makes them stressful for you?" The second instrument, the Children's Source of Stress Scale (CS), consisted of a 51-item self-reported checklist that helped students identify sources of stress. For this measure, Cronbach's alpha was from .89 to .93. The third instrument, the Children's Stress Symptom Scale (CSSS), was a 24-item self-report that asked students about stress related symptoms. The scale was from 0 (*once in a while*) to 2 (*most of the time*) for duration of symptoms. For severity, 0 was *not bad* and 2 was *terrible*. For CSSS, Cronbach's alpha was .99. Such a high level of reliability greatly increased the confidence that the scale measured the intended variables.

The fourth measure, Schoolager's Coping Strategies Inventory (SCSI), was a 25-item self-report scale measuring coping strategies used by children

from 8-12. The frequency of strategies used were rated from 0 (*never used*) to 3 (*most of the time*). For this measure, the frequency scale had a Cronbach's alpha from .75 to .82. Students reported about the effectiveness of strategies used, using a scale of 0 (*never do it*) and 3 (*helps a lot*), with a Cronbach's alpha of .79.

After data was gathered, researchers used single-group ANOVA repeated measures to determine changes in sources of stress, stress systems, and coping strategies. Researchers found that there was a significant difference in stressors over time. Students felt high stress during T1 and T2 (October to February), which then decreased throughout the rest of the year. Some significant stressors included sibling fights, missing family members, participation in sports, and taking tests. There were significant differences in the number of reported stress symptoms ($F = 26.008$, $p = .000$), which increased between T1 and T2 (October to February), decreased in T3 (March), and remained unchanged through T4 (posttest in April). Students reported hunger, headaches, and fatigue as top stress symptoms.

Skybo and Buck (2007) inferred that students adapted to new surroundings and the types of stressors decreased over time; researchers used the increase from T1 to T2 and then the subsequent decline as support for their inference. This study was strong in the area of outlining methods, and the Cronbach alphas for reliability measures were provided. One weakness of the study was that there was no racial diversity, which weakens credibility. Since the study was conducted in one school, there was limited transferability, unless the

demographics in future research were similar to the school being studied.

When schools use student tracking, some possible effects are on social comparisons, self-esteem, and students' tendency to ask for help across different achievement statuses. Butler (2008) used a quantitative, quasi-experimental study to examine if a relationship existed between the use of tracking in math and the variables of student ego orientation (self-concept about ability) and student help-seeking behavior. Butler hypothesized that the use of tracking would suppress the motivation for higher-achieving students because of the perception that asking for help might be "diagnostic of low ability" (p. 8). Butler also stated that she thought low achieving students would have higher occurrences of help-seeking in tracked classes because they were already perceived to be in classes of low ability. The researcher discussed the idea that students had a tendency toward, "demonstrating superior, or masking inferior ability," where the former was referred to as "approach" and the latter referred to as "avoidance" (Butler, 2008, p. 7).

The study involved 941 students, of whom 472 were boys and 469 were girls; the mean age of students was 11.4 years. Students were enrolled in elementary schools that either tracked only in math (n = 337 students) or did not track in any subject (n = 594 students). Thirty-two Grade 5 and Grade 6 math classes in 10 West Jerusalem schools were selected because of similarities in student demographics.

Butler administered the Motivational Orientations Questionnaire, using a 5-

point scale where 1 meant *strongly disagree* and 5 meant *strongly agree*. Students were asked to report about task, work avoidance, ego approach, and ego avoidance. The task referred to the actual math subject work at hand; work avoidance referred to students trying "to get by with little effort" (Butler, p. 7); and ego approach/avoidance referred to social comparisons students made with other classmates they perceived having different ability. Butler listed, "I got a higher grade than most other students" [ego approach] and "I didn't do worse than most other students" [ego avoidance] as examples of questionnaire items (p. 10). For adaptive help-seeking, students self-reported on a scale of 1 (*never*) to 5 (*always*) about different scenarios where they would seek out teacher assistance. Cronbach's alpha for the measure for adaptive help-seeking was .83, the mean was 3.73 and the standard deviation was .89. Between .8 and .9 is considered to be a good measure of reliability. The standard deviation indicates that the dispersion of scores was clustered more toward the mean of 3.73, with a smaller spread of variance across the larger sample.

Students answered a seven-item questionnaire related to math self-concept, which was adapted to work with math from Harter's Perceived Competence in Children Measure. The instrument used a range from 1 (low competence) to 4 (high competence); Cronbach's alpha was .86, the mean was 2.89, and standard deviation was .74, which coincided closely with the teacher-generated student ratings. Since Israeli schools did not rely heavily on standardized tests, teachers placed students in high, average, or low

achievement levels within the context of their classrooms. Teachers rated students from 1 (very low) to 7 (outstanding) based on performance related to the grade level math curriculum.

Within tracked schools, students were placed in one of three groups: low, average, or high track. In untracked schools, students were placed in groups according to teacher ratings. Tracking structures could have led to an increase in ego (ego orientation) and work avoidance goals across different ability levels. Higher-track students were dissuaded from seeking help in math class, while effects were minimal on lower-track students. Butler (2008) calculated a strong correlation between the measure used for math self-concept and teacher-reported ratings of achievement: in tracked schools, $r(337) = .50$, and in untracked schools, $r(594) = .52$. R-value is representative of Pearson's correlation coefficient. R values run from -1 to 1. Since Butler's values fell between 0.5 and 1.0, they were indicative of a strong positive correlation.

In tracked classes, Butler (2008) found no significance between task and ego approach, and a moderate correlation between ego approach and ego avoidance. High-achieving students engaged in ego-approach behaviors, where they highlighted superior academic ability. Low-achieving students engaged in ego-avoidance behaviors, where they avoided doing worse than others on different tasks. Boys scored higher than girls in all criteria, with the exception of task goals, where gender was considered to be an insignificant factor. The study had internal and external validity, in that it could be generalized for further study

within the United States. One necessary adaptation would be to pay attention to teacher-reported scale and heavy reliance on standardized test scores in the United States. The researcher also noted that in Israel, the use of student tracking was overt, whereas in the United States, the use was covert. The results were consistent and could be repeatable from one study to the next. However, it could prove challenging to find demographically comparable schools. In the study, Butler (2008) was able to examine schools that tracked students in math class only, and the untracked schools did not use instructional practices, other than mixed ability grouping, across subjects. One drawback is that the comparison of schools that tracked across multiple subjects with schools that did not track at all might yield different results. The study methodology is clear and has objectivity.

Hulan (2010) identified that there was a difference between student-led discussion and teacher-led discussion in guided reading groups. Hulan worked with a third-grade classroom that did not make adequate yearly progress (AYP), as defined by the No Child Left Behind legislation.  Hulan's observational case study led to the finding that when the teacher was absent, students usually continued discussion about the text, students asked more clarifying questions, and students offered more evaluations about answers than when the teacher was present and led the discussion. The teacher tended to dominate the conversation with funneling, or leading students to the specific right answers, rather than allowing students to elaborate or evaluate responses.

56

Three reading groups were included in the study. One group was considered to be at grade level, one was considered to be below grade level, and one was two years below grade level. The sample size was 24 students with 12 girls and 12 boys. The classroom teacher was in her second year and had been selected by the principal to take part in the study. The school had not made AYP, and Rigby Literacy K-3, a prescriptive reading series, had been adopted.

Students were placed in guided reading groups based on reading ability, leveled texts were selected, and the texts were aimed at developing specific reading strategies. In the classroom, the teacher led book clubs, and students led literature circles. Hulan (2010) used recorded and transcribed discussions, student surveys, and notes about teacher responses to researcher-generated questions. The researcher coded transcripts from the guided reading groups. Codes used included themes of reports, elaborations, and evaluations. Out of 653 coded units, 130 responses (20%) were coded by a trained assistant. The reported inter-rater reliability was 94% for coded responses, and researcher had responses that matched the independent coding 94% of the time in categories of themes of reports, connections, elaboration evaluations, and clarifications.

Of these responses, 561 occurred when the teacher was present. When the teacher was absent, students often continued to work with text discussions. While the teacher was present, students responded with reports (which required little, if any evaluation) 42% of the time, as compared to 17% of the time when the teacher was absent. Students asked questions or provided clarifications only

15% of the time when the teacher was present, but did so 29% of the time when the teacher was away. Hulan (2010) found that student engagement was very high when the teacher was absent. She also gave out a student survey in which students were asked about what they thought would help them become better readers. Student-generated responses included having the teacher call on them more, reading comics or other texts, and more free reading time in school. None of the students mentioned being interested in reading more Rigby books, while the teacher—in an informal interview—stated that students "love the Rigby books" (Hulan, 2010, p.60)

The small sample size of one classroom diminishes transferability. While the small sample size could be easier to replicate for future studies, the demographic breakdown, teacher experience, and type of instructional practice used could be more difficult to replicate. The data collection was transparent and thorough. The observational case study allowed the researcher to explore relationships within guided reading groups, teacher-student interactions, and student-student interactions.

Krebs and Roebers (2010) explored student self-awareness, or metacognitive strategies, used during test-taking. One hundred and seven students participated in the quantitative study; 47% (50 students) were female and 53% (57 students) were male. Fifty-four participants were eight or nine years old, while 53 were eleven or twelve years years old. Students were selected from six different public schools in Bern, Switzerland. First, students watched an

educational film, and then they engaged in small group interactions. During the first showing, the students were encouraged to pay close attention. During the second showing (one week later), students were told about the upcoming Cloze test. The Cloze test consisted of 26 questions; a researcher read aloud each question and students answered each question in writing. One researcher mentioned crossing out wrong answers as a useful strategy. Students were also asked to answer questions about Confidence Judgments (CJ). Cloze questions about the film were of two kinds: answerable (easy, medium, or difficult questions) and unanswerable. Students then indicated confidence judgments with every answer, and students had the option to withdraw their answers. Students received red pens and instructions to cross out answers that they perceived as being incorrect, a technique that Krebs and Roebers (2010) referred to as "free report."

Students had fairly adequate monitoring when determining if the answer was easy, difficult, or unanswerable. Krebs and Roebers (2010) found students ages eight and nine were able to correctly answer easy questions 70.4% (18.4 questions) of the time, and 11- and 12-year-olds were able to correctly answer easy questions 88.4% (13.2 questions) of the time. For all questions, younger students were able to correctly answer 50.3% (15.9 questions) of questions, and older students were able to correctly answer 64.6% (14.0 questions) of questions. Students also had fairly adequate monitoring when distinguishing between correct and incorrect answers. Researchers performed ANOVA on the

factors of correctness of answer, type of item difficulty, and student age: $F(1,76)$ = 19.89, $p < .001$, with an effect size of .21. The F-value indicated that the three variables were statistically significant when analyzed together, given the p-value of less than .05. For confidence judgments, students gave lower CJ to the incorrect answer. Older students were able to successfully implement the researcher-suggested strategy of crossing out the wrong answer. Younger students seemed to demonstrate attempts to use the strategy, but also ended up crossing out the correct answers. Researchers noted that the most important part of "strategic regulation of test accuracy" (p. 336) came from students' ability to withdraw potentially incorrect answers.

One drawback for the study is that students had to be fairly proficient in German in order to participate. This limitation could have created a sample that was not representative of the demographics of the area of study. The study had reliability, and researchers were able to find similar patterns between their findings and previous research findings. Krebs and Roebers' (2010) study could also be extended to include other age groups in different settings. However, for the results to be subject to comparison, researchers would need to align studies with the language skills, chronological age, and sex of participants included in the original study.

Crumpton and Gregory (2011) examined how intrinsic motivation was related to the factors of academic relevancy and student schooling outcome, and they used a sample size of 44 low-achieving ninth-grade students as part of their

longitudinal study. Half of the participants were men and half were women; 68% were African American and 21% were European American. Crumpton and Gregory defined *academic relevancy* as "the concept of students feeling that schoolwork is meaningful in their lives" (p. 42). Students were given a survey at the end of the 2006-2007 school year and again in the 2007-2008 school year. After completing each survey, students were interviewed.

Researchers tested for academic relevancy by using a 4-point Likert-type scale, where 1 meant *very true* and 4 meant *not true at all*. Cronbach's alpha coefficient for the academic relevancy scale was 0.86, and the internal consistency was alpha = 0.82. To test for academic achievement, researchers collected yearly GPA from student records. To determine academic engagement, survey data was collected, using a 4-point Likert scale, where 1 meant *not at all* and 4 meant *very much*. For this scale, Cronbach's alpha coefficient was 0.84 and internal consistency was alpha = 0.88. Researchers also gathered data to test for intrinsic motivation by using Harter's Scale of Intrinsic and Extrinsic Motivation (using only the intrinsic measures), which used a Likert 4-point scale, where 1 was *very true* and 4 was *not true at all*. For Harter's scale, Cronbach's alpha coefficient was .96 and internal consistency was alpha = .90. Because a Cronbach's alpha of greater than .90 is considered to have excellent reliability, confidence in the scale used increases.

Researchers found that students held a wide range of perspectives about school relevance and their lives. The mean was 2.54 with a standard deviation of

.73. Sixty-three percent (29 students) reported that the material was relevant in their lives, while 37% (17 students) reported that material was irrelevant. Students who participated in the study were considered to be low achieving, where their average grades fell within the D+ to C- range, where the mean was 1.89, standard deviation was .64; minimum value was 0.59 and maximum was 3.29.

Researchers chose participants who were at academic risk of failing out of school, and who were at risk of disengaging from the academic subject matter. Nearly all of the students (98%) reported their engagement in school as being between *sort of* and *very* engaged. The mean was 3.04, and the standard deviation was .53, with a minimum value of 1.82 and a maximum value of 4.00. Students also experienced a wide range of intrinsic motivation, from *not at all* to *very*. The mean was 2.46, and the standard deviation was .78, with a minimum value of 1.06 and a maximum value of 4.00.

Students who reported having higher academic relevancy also reported having higher intrinsic motivation, with higher levels of self-reported student engagement. Students who reported higher intrinsic motivation also reported having higher engagement during their second year, when compared with students who reported lower intrinsic motivation. Researchers found that there was no statistical significance in the correlation between GPA and variables.

The study had external validity, where results could be generalized to other settings. Research findings had objectivity, where the events that were

being studied were public and observable to others. The findings also suggest that higher relevancy was tied with higher motivation and engagement. Researchers found that there was no statistical significance between gender or race and academic performance. Crumpton and Gregory (2011) also found that students who reported having higher intrinsic motivation also reported having higher engagement during the second year of the study. Researchers concluded that neither relevance nor intrinsic motivation could be considered predictive of academic achievement.

Three studies are included in this section and researchers examined the way students perceived comparisons between themselves and other students. Do students make more frequent comparisons within-tracks or between-tracks? In tracked math classes, Chiu et al. (2008), reported 88% of participants made comparisons within their own tracked class while 12% of participants made comparisons with a different tracked class.

For ninth-grade students, Dumas, Huguet, Monteil, Rastoul, and Nezlek (2005) found that social comparisons could have influenced student self-concept of academic ability. In lower grade levels of fifth, sixth and combined seventh eighth, there was little correlation between the factor of social comparison and perception of ability. Researchers inferred that students in the ninth grade had spent longer time in tracked classes and could have made different associations of ability based on time spent in those classes.

Proponents for the use of student tracking argue that separate classrooms

allow gifted students to be sufficiently challenged. Bain and Bell (2004) discovered that at the elementary level, students who were identified as being gifted has different self-perceptions than students who were identified as being high achieving but not necessarily gifted.

Chiu et al. (2008) examined student use of within-track comparisons, across-track comparisons, gender, and self-concept in math. Researchers used data from 173 seventh-grade students from a school in Maryland. Students were 11 and 12 years old; 86 students were boys and 87 students were girls. The school was organized around three levels of math classes: Math 7 (below grade level), Pre-Algebra (grade level) and Algebra (above grade level). Sixty-six students were enrolled in Math 7, 43 students in Pre-Algebra, and 64 students in Algebra. The school used in the study outperformed the state in both math and reading, according to the standardized Maryland School Assessment (MSA). The participant school had shown a pattern of higher percentages of proficient or advanced students.

Chiu et al. (2008) wanted to know how the use of tracking could potentially affect self-concept and self-esteem; self-concept was related to content-area knowledge, while self-esteem was related to overall assessment of self-value. Chiu et al. also focused on social comparison, and included upward social comparison (when students identify with someone who is doing better than they are) and downward social comparison (when students identify with someone else as achieving less or doing worse). Researchers also noted that students could

also create comparisons within the same track and between different tracks.

Students were given surveys which were based on three factors: (a) self-esteem, (b) self-concept, and (c) social comparisons (within-class or between classes). Students had 45 minutes to complete the survey at the beginning of the year. After the data was gathered, researchers assigned student data an alphanumeric code to maintain anonymity. Harter's Self-Perception Profile for Adolescents was given, and specific scales of Global Self-Worth and Scholastic Competence were emphasized from that measure.

Students were asked to assess where they thought they were in math, and where they were subsequently ranked. Then students were asked to assess how they thought they performed in math, when compared with classmates and those outside of the class. Twenty out of 171 students (11.7%) considered themselves to be making comparisons outside of their math class or between math classes. A much larger group, 151 students out of 171 (88.3%), considered themselves to be making comparisons within their math track.

ANOVAs were performed; independent variables were the three track levels and the two levels of gender. The dependent variables were students' math self-concept, school self-concept, and self-esteem. ANOVA indicated that math self-concept was largely affected by math track, $F(2,166) = 3.091$, $p = .048$. For school self-concept, math track largely affected the concept, $F(2,161) = 12.011$, $p = .000$. Math track placement was not significant with regard to gender or self-esteem. One interesting finding was that when grades were statistically

controlled for (meaning they were not really given), effects associated with tracking disappeared. Students mentioned that it was easier to compare within tracks because they were all studying the same material, with the same teacher, at the same time. Comparing across tracks would increase the number of factors that would have to be taken into account.

Statistical analysis lent strength to the study's descriptive data. The research methodology was complete, thorough and clearly outlined, and the data was fully accessible. While the sample size was large, the study took place within one school and in one geographic area.  As a result, generalizations across different geographic areas and student populations may be difficult.

Dumas et al. (2005) explored the idea of the potential use of upward social comparison by elementary grade students. A total of 339 students in Grades 5-9 in French schools answered teacher-administered questionnaires, which asked students to rate their academic performance relative to their classmates in three courses (Math, Reading and Writing), to list two students from their class that they compared exam grades with, and to rate how important each academic domain was to them.

Teachers administered questionnaires during the last two weeks of Trimester 2. Students were asked to list two students from their class (referred to as Choice 1 and Choice 2) whose exam scores students used for comparison with their own scores. Other measures included how much students talked to the person whom they chose as Choice 1; at this age, it was common for those who

interacted frequently with Choice 1 to consider them good friends. Students also rated their academic performance   relative to the rest of their classmates within each course, using a five-point scale where 1 meant *much worse*, and 5 meant *much better*.

Students answered questions about how important each academic domain was. The pre-test indicated that the 5-point scale was being misunderstood, so it was narrowed down to a 2-point scale that included 0, which meant *not important to me*, and 1, which meant *important to me*. Students were able to leave spaces blank if they were unsure of how to answer. To determine grades, reports were gathered for each of the three courses. In the French education system, students were given grades on a scale from 0 to 20. Researchers reported the breakdown of grades was as follows: extremely poor (0-5), poor (6-9), passable to satisfying (10-14), and very satisfying to excellent (15-20).

For fifth-grade students, when students nominated others for comparison, t-tests confirmed that they did not choose students who were doing better than them in different courses (reading, writing, and math). Dumas et al. (2005) found that gender played a lesser role than previously thought, though it was statistically significant (p < .001) for Choice 1. Researchers determined that the greatest significance was found in the ninth-grade students, with regard to comparison and exam grades, when compared to the relationship between those two variables across other grade levels. In lower grade levels, a high percentage (66% for Choice 1 and 54% for Choice 2) of participants had same-sex

comparison targets. There was no support for the idea that there was a causal relationship between social comparison and resulting self-concepts of ability.

The study had credibility because the method was sound and the way in which it was conducted was also sound. The study structures were transferable, where the questionnaires could be carried out across different contextual settings. Low dependability was based on the fact that not a lot of previous research had been conducted on the topic of social comparisons in elementary school.  Although the study was conducted outside of the United States, if a comparable grade conversion could be determined, a similar study could be feasible within the United States.

Bain and Bell (2004) used a quantitative study to examine if gifted students in the fourth, fifth, and sixth grades had different social self-concepts, social attributions, and peer relationships, than students who were identified as being high achievers and who were not identified as being gifted. Researchers used a Self-Description Questionnaire (SDQ-1), which measured eight components of pre-adolescent self-concept behavior. Limited research had been conducted about social self-concept for gifted students, and the results of the studies had been conflicting (Bain & Bell, 2004, p. 170). Bain and Bell also used the Student Social Attribution Scale (SSAS); when Bell used this scale in a 1995 study, the "test-retest reliabilities for the internal scales ranged from .66 to .77" (Bain & Bell, 2004, p. 171). A test-retest value of .7 or higher was considered a good reliability, according to Cronbach alpha values. Test-retest demonstrated

repeated trials under the same conditions, using the same measure. Bain and Bell expanded the measure to include 30 items, and included subscales of Success Chance (SC), Success Task Difficulty (ST), Failure Chance (FC), and Failure Task Difficulty (FT). Cronbach's alpha for the expanded measures instrument constituted a range of .76 to .93. Bain and Bell had higher reliability based on the revised SSAS. SSAS examined the causal attributions of success or failure in socially related school situations. SSAS rated the 30 causal statements on a Likert scale, where 1 was *seldom* and 3 was *often*.

Bain and Bell (2004) administered scales to small groups consisting of 6-10 students. In addition, the teacher was given a measure called the Teacher's Social Rating (TSR), which asked the teacher to give feedback about students' peer relationships. TSR was modified to exclude identification of popularity status among students. TSR was based on a scale from 1 (high frequency of positive interactions with others) to 5 (high frequency of avoidance of interactions with others). Researchers found that TSR had construct validity, as correlation data supported TSR as a measure of peer relationships; TSR shared a significant correlation with the SDQ-I measure of peer relations ($r = -.23$, $p < .05$), while sharing no other significant correlations with other SDQ-I subscale measures.

Ninety-three elementary and middle school students took part in this study. Twenty-nine students were in fourth grade (approximately 31%), 32 in fifth grade (approximately 34%), and 32 in sixth grade (approximately 34%). The average age for students was 11.06 years. Only one student was Asian

American, two were African American, and 90 were European American. A possible threat to external validity for the study came from the lack of diversity found in the sample population, which decreased the generalizability of the study across different contextual settings. Twenty-six of 93 students (approximately 30%), were identified as being gifted and were given special courses in conjunction with their regular academic program. To be considered gifted, students had to be two standard deviations above the mean on an intelligence test and have academic achievement on or above the 96% percentile in subjects of reading, language, math, science, or social studies. Sixty-seven participants were in the 85th or higher percentile on a Terra Nova Comprehensive Test of Basic Skills.

Researchers found that there was no significant difference between gifted and high-achievers when variables of gender and self-concept were measured. In addition, interaction was not a significant variable. For gifted students, parental relations was not a significant variable. The mean TSR score for gifted students was 1.96, with a standard deviation of 1.09. The mean TSR score for high-achievers was 1.88, with a standard deviation of 1.4.

Gifted students scored higher in three quarters of the socially related self-concepts, and displayed stronger levels of attributions of social success being tied to the internal factor of ability.  Even though little previous research had been conducted with regard to gifted students and self-concept, Bain and Bell (2004) reported findings that supported other research theories. Gifted students also

had a higher rating associated between effort (M = 39.58) and ability (M = 35.15). Teachers rated male students higher than female students. This study posed an interesting research question about the comparisons being made between high-achieving and identified gifted students. How do they perceive themselves? How do they compare themselves to others? To what do they attribute their success? One weakness of the study design was that there was no history about past academic performance provided about the groups. There was a mortality rate of two students. A threat to external validity was the lack of racial and ethnic diversity, due to an oversampling of European American students. This raises an important question: Are 90% of the gifted population and high-achieving student population European American? How much diversity exists in these programs and in identifying students for these programs?

## Outside Factors

This section addresses factors including parental warmth and student monitoring, which are not directly addressed by the use of student tracking. However, these are important factors that have impacts on students' lives and which could help or hinder academic performance. Educators should understand factors that occur outside of school, and how those factors may manifest themselves through student behavior. Student tracking may not take into account the complexity of students' lives and experiences. This section addresses outside factors that could affect student performance within the classroom. Examination of outside factors helps frame the issue of student tracking within a

larger, real-world, context.

Students are influential from a young age. Patrick, Schrepferman, Snyder, and Snyder (2005) studied the relationship between factors including parental warmth and monitoring and student conduct problems. Researchers found when parents had higher levels of student monitoring in early elementary years, students had lower reported levels of conduct problems in later elementary years.

Student tracking can occur between classes, and can also occur between schools.  Gentrification could explain why some schools are considered to be lower achieving and others are considered to be higher achieving. DeSena and Ansalone (2009) examined gentrification and schooling inequity. While researchers used informal interviews to collect data, their findings included parent statements about preference to change schools rather than trying initiate change in local schools. Student mobility is an outside factor that could affect an institution's tracked status. Similar to DeSena and Ansalone's conclusions that gentrification can lead to social inequity in schools, Thompson, Meyers, and Oshima (2011) reported there were correlations between achievement in language arts and math and student mobility. School poverty level and student mobility rate also had a negative statistically significant relationship.

Could the type of school have an impact on student performance? Lopata, Wallace, and Finn (2005) investigated Montessori programs which promoted multi-aged classrooms and individual developmentally appropriate tasks for

students. Researchers found conflicting results as to the efficacy of Montessori programs when compared with other school formats.

Patrick et al. (2005) investigated whether the factors of parental warmth, communication, and tracking had an effect on early elementary conduct problems and later elementary self-monitoring during their nine-year longitudinal study. Researchers focused on 267 students: 29% were self-identified as African American, Hispanic/Latino or other minority status. Researchers defined early elementary to mean an average age 5.5 years and later elementary school to mean an average age 9.5 years. Missing data estimates were calculated using Amos 4.0, a program that used "full- information maximum likelihood (FIML) estimation method" (p. 1002).

Patrick et al. (2005) hypothesized that there would a positive correlation between parental warmth, communication, and tracking of students in early years, and monitoring of students in subsequent years. The main correlation researchers found was between the variable of monitoring and conduct problems. Child covert conduct problems were negatively associated with third- or fourth-grade monitoring, meaning that parents monitored students more often (b = -.30, and p < .05). Student growth in covert conduct problems was also negatively associated with an increase in monitoring during later years (b = -.51, and p < 0.05). Students with lower-quality parent-child relationships often had higher covert and/or overt conduct problems during the early elementary years, which resulted in a decrease in parental monitoring in later elementary years.

Patrick et al. (2005) concluded that familial and social factors were related to occurrence of student conduct problems, where closer parental monitoring during the early elementary years potentially led to lower conduct problems in later years. Worthy (2010), a study that will be reviewed in the next section, found that honors class teachers reported fewer student conduct problems than regular class teachers. Teachers perceived conduct problems as being related to lower student ability. Patrick et al. (2005) added the need for further research about the development of behavior in children and the corresponding quality and type of interactions children and parents were engaged with. The study had strength because the nine-year observation period helped account for variance. While the researchers relied on descriptive data, including videotaped conferences of parent-child interactions, they provided reliability measures for scales used and included inter-rater reliability from the coding process. For example, Patrick et al., (2005) used Achenbach's Child Behavior Checklist (CBCL) to measure overt conduct problems, and the reliability for the scale was .90. Patrick et al. were transparent about coding and reliability measures, making replication of similar conditions more plausible for future researchers.

DeSena and Ansalone (2009) explored the relationship between gentrification, schooling, and social inequity. Researchers focused on one neighborhood in Brooklyn, New York. In 2001, Greenpoint-Williamsburg had 160,338 residents: 48% were European American, 6% were African American, 38% were Hispanic, 4% were Asian and Pacific Islander, and 5% were self

reported as other (non Hispanic). Researchers used a case study format where participant observation was used in a preschool program sponsored by the YMCA. Researchers chose the study site because one researcher had a son in attendance in the program. Later, DeSena and Ansalone (2009) conducted formal interviews with 20 parents about factors including residential histories, educational levels, occupation, and decisions about schools.

DeSena and Ansalone (2009) defined gentrification as "the conversion of socially marginal and working class areas of the central city to more affluent residential use" (p. 62). Researchers addressed three different perspectives about the use of student tracking. First, tracking was the best way to properly select human capital and was viewed as being most efficient. Second, tracking created an increase in both achievement and self-esteem, where less competition was necessary between perceived similar ability group members. Third, access to learning was limited by the use of tracking, where low-SES students and minorities are often overrepresented. In the study, gentrification was related to between-school tracking (between separate schools), not between-class tracking (between separate classes housed in the same school).

In New York, charter schools were usually available for families of school-aged children. Under the administration of New York City's Mayor Bloomberg, students had been forced to attend neighborhood schools based on location proximity. Entrance into choice schools occurred in one of three ways: by application, through student testing, or through random lottery. Parents were able

to request waivers based on geographic boundaries, and siblings of students were given higher consideration for admissions than others, thereby decreasing the odds for other applicants. Parents tended to prefer charter schools, and one parent openly criticized one of the neighborhood schools as teaching only to the test. Families decided to enroll students in the Talented and Gifted (TAG) programs in Manhattan in order to sidestep dissatisfaction with local schools. DeSena and Ansalone (2009) discovered a school placement loophole, where placement in kindergarten guaranteed placement for the next 6-8 years, and kindergarten was not mandatory in New York. Families chose TAG programs because the other alternative for these families was homeschooling.

Parents reported to DeSena and Ansalone (2009) that it was easier to transport students to different schools than to instigate change in local schools. Another significant finding was that the major consequence of gentrification was an increase in stratification between social classes. One parent considered moving the family outside of the neighborhood until after their child had been accepted. Another parent spoke out about the local schools not having extra time to include her when she visited schools.

DeSena and Ansalone (2009) conducted a qualitative case study, but  the study sample size (20 participants) was very small and not random (a sample of convenience), both of which decreased the dependability and credibility of the study. The study lacked dependability to the extent that limited research had been done on the effects of gentrification on attendance at local schools. The

study had lower credibility because there was no mention of coding answers from the informal or formal interviews, and there was very little reference to actual answers given. DeSena and Ansalone's (2009) discussion and conclusions represented a synopsis of their ethnographic research, but were not clearly data-driven.

Thompson et al. (2011) studied the effects of student mobility on schools' adequate yearly progress (AYP). Researchers used a quantitative, correlation design, and used correlation analysis to examine the possible relationship between student mobility (movement from one school to another) and achievement. Since AYP was crucial to access to federal funding, researchers wanted to examine if student mobility had an effect on AYP, a set of statistical measures required by the No Child Left Behind legislation. Researchers used regression analysis to determine if student mobility had an impact on different levels of achievement, when the variables of school size and poverty status were considered. Among the 1,062 participating schools were some that had made AYP and schools that had not. The schools that did not meet AYP were placed on "corrective action status", as reported by the Government Accountability Office (GAO) in 2007.

Data was retrieved from the state educational agency and the academic competency test, which the researchers referred to as ACT in the study. ACT was used to measure reading, language arts, and math by grade level. The sample size included all elementary schools from a southeastern state.

Composition included urban and suburban schools. Schools that were excluded did not have the requisite kindergarten through fifth-grade structure for comparison; pyschoeducational centers or those that served smaller levels of different grades were removed from the study also. Of the 1,062 schools, 971 had met AYP (including 633 Title 1 schools) and 91 had not met AYP. Of the total, 717 schools were considered high poverty.

Thompson et al. found that the correlation between social mobility and reading achievement between grade levels ranged from r = -.54 to r = -.61. This coincided with the correlation Thompson et. al. found between school mobility and language arts (p < .001, r = -.46, to r = -.55), and between school mobility and math achievement (p < .001, r = -50, r = -.59).

Researchers found no statistical significance between mobility rate and school size (r = -56). There was statistical significance between mobility rate and school poverty (r = -.434, p < .001). When researchers controlled for size and poverty status, mobility rate and ACT was significant (p < .001) across first through fifth grades.

Findings also suggested that schools that had met AYP had a lower rate of mobility (m = .23, and standard deviation = .09). The same statistics for schools that had not met AYP were m = .28, with a standard deviation = 0.08. For schools that had met AYP, school poverty and mobility had a stronger correlation (r = -.429, and p < .001). For schools that had not met AYP, school poverty and mobility had a weaker correlation (r = -.225, p = .032). Thompson et.

al. (2011) determined that out of the three subject areas—reading, language arts, and math—reading had the strongest negative correlation with school mobility.

While this study had findings that were supported by previous studies, it also had findings that conflicted with past studies. This is a multifaceted topic that may have greater complexity than the researchers can control for, indicating that further research may be justified. A weakness of this study is that there was no mention of how statistics were applied to data (including if statistical software was used), and only the results of analysis were provided. Reliability was decreased by the omission of statistical measures, which could have decreased predictability of results across future studies. Thompson et al. (2011) did not provide information about the number of individual students involved. Finally, there was an oversampling of schools that had met AYP, which may have had some significance on the results or findings. One study strength was the very large size of the study itself, which included 1,062 schools, or all elementary schools within a undisclosed southeastern state; variance based on random occurrence would have been decreased due to the large sample size.

Lopata et al. (2005) identified Montessori schools for study because of multi-age classrooms, the dual roles of teacher as observer and facilitator, and claims that attendance at Montessori schools could lead to higher academic achievement. The original aim of Montessori schools was for low income and students with developmental disabilities to be able to learn in an environment that would be conducive to their specific learning needs. Montessori teachers used

developmentally appropriate and individualized learning strategies, as opposed to strategies that forced students to learn at a specific pace.

The introduction of federal funding for magnet programs helped increase the number of Montessori schools. More schools opened up as student-directed learning gained credibility. Montessori students were also encouraged to choose their own activities, increase their use of the manipulation of objects to facilitate learning, and engage in collaborative problem-solving with minimal teacher intervention. Montessori schools used strategies that starkly contrasted with those at more traditional schools.

Lopata et al. (2005) conducted a quantitative, correlation study. MANCOVA was used to analyze variables across Montessori schools, open magnet schools, structured magnet schools, and traditional non-magnet schools. Gender, ethnicity, and socioeconomic status were the main variables that were controlled for during analysis.

The study included 543 students, of whom 291 were in fourth grade and 252 in eighth grade; all were from an urban district in western New York state. Four schools were chosen to participate, based on similar school profiles, as reported by the New York State Education Department. Lopata et al. (2005) matched the schools on variables of gender, ethnicity and SES. SES was determined by the federal formula for free and reduced lunch. Sixty-seven percent of students were identified as low income (69% of fourth grade and 64% of eighth grade). The ethnic composition was 47% European American. The rest

of the student population, who reported themselves as African American, Hispanic and other, were considered to be part of a category of "minority" for the purposes of this study.

Two standardized test results were used: the New York State Mathematics and English/Language Arts (ELA) exams and the math and language art portions of the McGraw-Hill's 2002 Terra Nova exams. Data was provided by the school districts. Student data was compiled anonymously, with the use of district-assigned numbers and without the use of personal information.

Lopata et al. (2005) reported that gender was not related to math achievement at either fourth- or eighth-grade level across any of the schools. Gender was related to language arts at both grade levels; female students scored higher in both grades. Both ethnicity and SES were significantly related to subject achievement. Minority and low-income students had lower scores in both subjects across grade levels, compared to European American students who were not identified as being low-income. After performing MANCOVA, Lopata et al. (2005) found that the three factors of gender, ethnicity, and SES were significantly related to achievement: Grade 4, $F(6, 546) = 19.0$, $p < .001$; Grade 8, $F(6, 476) = 13.9$, $p < .001$.

In the fourth grade, there was no significant difference in math between Montessori and Structured Magnet (SM) schools; Montessori students had higher achievement than Open Magnet (OM) schools by .60 standard deviations. Montessori students had lower math than traditional non-magnet (TNM) schools,

and the effect size was .37 standard deviations. No difference was found with regard to ELA among the schools being examined.

In the eighth grade, Montessori schools had lower language arts achievement than SM and TNM schools (the effect size = .77, SD = .59). Montessori schools also had lower language arts achievement than OM schools, but the researchers found the results to be statistically insignificant ($p < .06$). No difference was found with regard to math and the schools being examined. An interesting finding is that in fourth grade there were larger differences in math than in ELA, but these results were reversed in the eighth grade.

Researchers found conflicting results about the effectiveness of Montessori schools in comparison to other school formats. Lopata et al. (2005) hypothesized that Montessori students would perform academically better than students in Traditional Non-Magnet schools and Magnet schools alike, in the subjects of ELA and math. Researchers could not support their hypothesis due to mixed and conflicting findings. The study had external validity, in that it could be generalized across other settings. One drawback to the study was the high proportion of European American students in relation to the so-called "Minority" students. Researchers did not make specific distinctions about the unique student populations involved in the study. As a result, findings could conflict if the study was generalized to more diverse student populations. Previous research studies have had conflicting and mixed accounts of the effectiveness of Montessori schooling as compared to non-Montessori schooling. Therefore, the

82

reliability of the study was decreased, because the results could not have been considered predictable when replicated across future studies. The study had objectivity, because it was conducted in a way that was both public and observable to outside investigators. The Lopata et al. (2005) correlation study had decreased internal reliability due to the fact that they could not establish that variations in the school format (independent variable) were related to variations in student academic performance (dependent variable).

## Positive Results from Tracking Structures

This section addresses studies which show that ability grouping (mentioned more than tracking) is not in and of itself harmful. Instructional practices that accompany ability grouping can support or hinder the efficacy of its use. This section starts to address why the use of ability grouping could be beneficial for students.

Ability grouping can yield positive academic gains. Educators should be intentional in their groupings. Groups should be flexible, meaning that students are not permanently placed in specific groups, and different groups should be used for instruction across specific content areas. McCoach et al. (2006) found that when teachers used ability groups in reading more frequently, the school experienced an overall gain in reading. Researchers concluded that the improvement in reading performance could have been attributed to the use of flexible reading groups. Lyle (1999) found similar results when studying the use of mixed ability groups, where students showed academic improvement in

reading and writing. While McCoach et al. studied homogenous ability grouping, Lyle investigated the effects of heterogeneous ability grouping. Proponents for the use of tracking argued that higher achieving students are being insufficiently challenged when placed in heterogeneous groups. Lyle's results imply that the use of mixed-ability groups could be beneficial for all students involved. Taken together, the two studies indicate that ability grouping is not necessarily a poor instructional choice. If used strategically, students could stand to gain from the use of such instructional groups.

McCoach et al. (2006) studied the possible effects of ability grouping on reading gains during a kindergarten school year. The researchers used data from the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K), which had been gathered by the National Center for Education Statistics (NCES) in 2000. The NCES had the goal of creating a nationally representative sample of United States kindergarten students from the 1998-1999 school year. According to the NCES data, Asian-Pacific Islander (API) students were oversampled 2.5 to 1, where there were 2.5 times more API students sampled than were representative of the U.S. population at the time. McCoach et al. used a model-based approach derived from Korn and Graubard to help diminish influence that may have resulted from oversampling.

The NCES used an item response theory (IRT) scaled cognitive assessment during data collection. Theta, an "ability estimate" (McCoach et al., 2006, p.341) was derived from the reading assessment scores. Theta was used

to make student comparisons and to measure time-lapsed achievement gains. In this study, researchers subtracted fall IRT reading scores from spring IRT reading scores to document growth.

McCoach et al. (2006) explored the potential relationship between the use of within-class ability grouping and reading gains. Teachers homogeneously grouped students based on ability within their larger, heterogeneous classroom. Researchers hypothesized that there was a positive relationship between reading gains and frequency of use of within-class ability grouping. The researchers used multilevel analysis and controlled for child-level and school-level variables. Child-level variables included reading gain, proportion of boys, fall reading IRT scores, socioeconomic status (SES), and assessment age (measured in months). School-level variables included frequency of use of ability groups, proportion of full-day to half-day kindergartens, and school aggregate data for reading gains and SES. The sample size included 10,191 kindergarten students, data was taken from 620 schools, and the average number of students per school included in the sample was 17.7; a standard deviation of 3.97 for the sample size per school existed. McCoach et al. used a Hierarchical Linear and Nonlinear Model (HLM) Version 5.0 for analysis. Researchers also included the means for fall reading score and SES, which served as Level 1 predictors, during the analysis of Level 2.

McCoach et al. (2006) determined that reading gain occurred across all individual subjects and that distribution of gain scores was positively skewed. Out

85

of 10,191 students sampled, 2.3% (approximately 230 students) had zero or negative gain in reading during kindergarten. The range of change in reading was from -21.44 to 43.63, the mean was 10.20, and the standard deviation was 6.11. Teachers self-reported a mean of the use of within-class ability grouping of approximately once per week, where 30% of teachers reported not using within-class ability grouping, and 44% of teachers who did use within-class ability grouping reported using this grouping strategy for 16- to 30-minute increments during the day. Researchers found that higher frequencies of use of ability groups were positively correlated with an increase in overall school mean reading score. Researchers concluded that the frequent use of ability groups could have contributed to the improvement in early literacy and reading scores.

In addition, researchers reported conflicting results that they inferred were due to school-level variables that were not examined in their study. McCoach et al., determined there was a statistically significant, negative relationship between initial reading scores and reading gain ($p < .001$). Researchers concluded that the higher the initial reading scores, the lower the overall reading gain. Researchers recommended that flexible ability groups be used as part of a differentiated instruction curriculum.

The study has limitations. One threat to external validity is that identified limited-English proficient students and identified learning disabled students were excluded from the study sample. This is a threat to reliability because it excludes students who are naturally included in general education kindergarten

classrooms.  Another threat to external validity is that student mobility was not included as a factor in analysis. Student mobility is indicated when students transfer or move from one school to another. Student mobility is a reality of contemporary schools, and may not be feasibly controlled for if this study was conducted in a different setting. The study sample was further restricted by including schools that had five or more students in the ECLS-K data sample. McCoach et al. (2006) did not provide the number of teachers in the sample being studied.

Further research would be needed before concluding that this was a representative sample of individual student reading gains as a result of within-class ability grouping. No data was included about the specifics of individual kindergarten students, although aggregate school data was provided.

Researcher Lyle (1999) used qualitative interviews of two groups of 12 students to examine the question of how the students felt about the use of mixed-ability grouping to enhance literacy. Twelve research assistants collected data from the 10-week project, and recorded data from one afternoon per week (10 during the collection period)). Students were 9 and 10 years old, all interviews were videotaped and then transcribed, and all interviews were conducted separately. The first group, designated Group 1, was below-average attainment. The second group, designated Group 2, was above-average attainment. Teacher-supplied information and Suffolk Reading Test results were used to create the different groups. Ten weeks of group work and interactions were

observed. After Groups 1 and 2 were established, students were then placed in mixed ability groups; groups were mixed in gender and specifications about how groupings occurred were not explicitly stated. When interviewing students, Lyle chose to reestablish the initial Groups 1 and 2 and report their results separately.

During the first week, students were asked to focus on the Nigerian village of Gashaka and make predictions of about factors of cultural values, norms, and environment. A culturegram, or poster collection of photos and illustrations, was also examined by students. Finally, the teacher read aloud a story about a 10-year-old boy from a Gashakan village.

During the second and third weeks, the students worked together with a card-sorting activity. Students were observed to determine if they were able to integrate the new information introduced by the cards with the previous information they had learned.

Students were given high-level cognitive demand reading in the fourth week about animals that live in the Gashaka-Gumpti animal reserve. Lyle (1999) noted, "The reading is demanding, mirroring the expectations of encyclopaedias or reference books" (p. 286). Students were assigned the task of creating a database for the animals and providing summaries of important information. Students wrote and illustrated a picture book about Gashaka during week five. Students were reminded to remain aware of audience and purpose while composing their writing.

During weeks 6 and 7, students were regrouped into larger groups of eight

students and asked to role play. The students created a three-minute clip for their fictional news show called "Newsround". In weeks 8 and 9, students made a board game about the types of problems that game-reserve guards were facing. Teacher direction included a short "design brief" to help determine the purpose of the game, but all other aspects were left up to students. Students focused on audience, prior knowledge, and game construction. In week 10, students made an informational brochure to attract more visitors to the Gashaka-Gumpti national park.

One drawback of the study is that it does not list when the interviews were conducted, or the intervals during the 10 weeks in which the interviews occurred. There are fascinating classroom implications here, that units can be interesting and engaging, and that curriculum can draw upon students' interests. This study demonstrated how much planning has to go into creating a unit like this, and how keeping the level of cognitive demand the same for different students was very important; everyone had to feel a strong sense of personal agency, with something to contribute to the groups.

Group 1 (below-average attainment) students explained that having peer assistance with tricky and unfamiliar words was helpful). All Group 1 students felt that the group reading was what they benefited from most, and that they were all able to share ideas. For writing, students in Group 1 reported that note-taking was helpful for them while reading. Group 1 students also discussed how important peer-to-peer support was for them, because it aided in mutual learning

during participation in the educational activities. Group 1 members noted that their collective group work skills had increased, because students had experienced the value of what happened during discussions when disagreements were present.

Group 2 (above-average attainment) students reported that the strategy used by some Group 1 members, phonetically sounding out words, was not necessarily a successful strategy. Group 2 members stated that the ability to say the word was not equal to the ability to define or understand the meaning of the word. In their mixed-ability groups, Group 2 students were able to gain more familiarity with words, which they reported helped them with their own reading and decoding. Group 2 members began to make a connection between the help they were giving the Group 1 students and their own knowledge of resources and words. Students in Group 2 noted that writing for different audiences was important. Lyle (1999) determined that most of the answers to the classroom discussions about what students had learned were both detail-oriented and laden with examples; which held especially true for female students (p. 292). Group 2 students also reported that they learned group skills, including democratic decision-making and increased friendships.

The results indicate that both groups benefited from the mixed-ability grouping and the reliance on collaborative activity during literacy was critical to their academic success for the task. One important part of Lyle's (1999) study was that it emphasized the use of student discourse about literacy skills. The

study included no geographic area, school information, or demographic information about students. No specific reference was made about the methods of transcription of data. The study had little credibility, transferability, confirmability, and dependability. The researcher could have increased the strength of the study by providing more information about the subjects and context.

## Negative Results from Tracking Structures

Two studies clearly indicated that ability grouping or tracking did not work. While previously reviewed studies in this paper were concerned with alternative practices or self concepts of students within the grouping, these two studies demonstrated what people might imagine when they think of tracking. One study is included in this section because the overall findings were negative, but there were mixed results about the use of ability grouping.

Student tracking is used and perceived differently in the U.S. than in other countries that use similar practices. LeTendre, Hofer, & Shimizu (2003) compared the U.S., Germany, and Japan and conducted interviews with school personnel, parents, and students. In Germany and Japan, tracking was a known practice, whereas in the U.S., tracking was a lesser known practice. In all three countries, track placement was determined by student language acquisition rate and lower tracked classes consisted of higher concentrations of lower SES students. Is ability grouping a new name for student tracking? Worthy (2010) studied honors and regular language art classes in Texas middle schools.

Worthy reported that teachers exhibited different behavior towards students in different level classes and expressed different expectations for student behavior in different level classes. Worthy concluded that the detrimental labels being applied to students was more pervasive in contemporary ability groups than in tracked classes of years past.

One study with mixed findings about ability grouping was included in this section. Due to the fact that overall, researchers concluded that ability grouping was detrimental, the study was included under the heading of Negative Results from Ability Grouping. Sørenson and Hallinan (1986) determined that students in grouped classes were taught less material than students in ungrouped classes. In addition, researchers reported that students in grouped classes might have learned more in grouped classes because of instructional methods used to fit the needs of specific classes. Sørenson and Hallinan's study may have influenced the research questions used by LeTendre et al., and Worthy.

LeTendre et al. (2003) examined what the term *tracking* meant in the United States, Germany, and Japan. How do other countries define student tracking? Researchers used data from the Trends in International Mathematics and Science Study (TIMSS) case study project. Study participants included teachers, parents, and students; they were selected by consultation with local school personnel. In Germany, 366 hours were spent gathering interview data from 199 parents, students, principals, assistant principals, and counselors. In Japan, 494 hours were spent interviewing 247 similar participants. In the United

States, 542 hours were spent interviewing 271 participants.

Researchers used 250 hours of observations in science and math classes within the three countries as the basis of their qualitative study. To increase validity, researchers discussed the merits of different themes, reflected on conclusions, and presented findings to colleagues for further review and revision.

Tracking is the practice of differentiating students on the basis of real or perceived differences in ability. Researchers concluded that the use of tracking in all three countries could hide negative impacts on underperforming ethnic, racial, and low-SES student populations. LeTendre et al. (2003) reported that language acquisition rate was directly related to track placement. For example, if students acquired language at a slower rate, they were often placed in lower track classes.

Researchers used qualitative data for their research, and statistical analysis was not applied. The LeTendre et al. (2003) study was a comparison across three countries and was based on phenomenology, in an attempt to subjectively approach objective measures, such as personally held beliefs about the use of student tracking. Because of the unique parameters of the study, LeTendre et al. (2003) did not have findings that were consistent with previous studies. Since researchers were analyzing secondary data (from the TIMSS case study information), there was little specific information about study participants; study credibility and transferability were decreased by absence of necessary demographic information.

LeTendre et al. (2003) found that student tracking was based on real or perceived differences in student ability. All three countries used tracking; Germany and Japan were explicit (overt) about using tracking, but the United States was implicit (covert) about using student tracking. Researchers speculated about whether the explicit versus implicit use had to do with cultural differences and resulting attitudes about the use of tracking. Researchers also found that the negative impact that tracking had on minority groups, immigrants, and families with lower socioeconomic status (SES) were comparable across countries.

Worthy (2010) studied ability grouping and was interested in determining if ability grouping was another name for the practice of student tracking. Worthy focused on 25 sixth-grade teachers, each of whom taught at least one honors and one grade-level (regular) English class, selected from eight urban Texas middle schools. Worthy conducted interviews with a recurring theme of the use of ability grouping. Worthy found that teachers drew clear delineations around entire groups of students, that instructional methods differed between grade level and honors courses, and that teachers held negative perceptions of underachieving students which were demonstrated through words and actions.

Worthy's (2010) qualitative study started as an interest in how language arts teachers engaged in the act of reading instruction. However, the researcher found that clear differentiation was made between honors middle school language art classes and regular middle school language art classes, so the

driving research question was changed. Teachers made distinctions based on parental involvement, student behavior, and student ability between the different levels of classes. One teacher described honors students as being "every teacher's dream" (p. 272), while another teacher described regular students as being "very very low. They are my most difficult class, and my biggest discipline problems are in that class" (p. 272).

Teachers of both honors and regular courses relied on perceptions about achievement, readiness, and expectation as indicators of ability. Students were discussed in the context of general categories, and individuality was not emphasized. During the observation phase, the researcher observed stark differences between the instructional methods used for each class level. Teachers used transmission-based instruction in regular classes, while teachers used creative or transaction-based instruction in honors classes. Teachers also made instructional choices based on their perceptions. For regular classes, teachers relied on basal readers and worksheets, and they assigned less silent sustained reading than in honors courses.

In addition, Worthy's (2010) findings indicated that negative perceptions of lower-achieving students were demonstrated both in words and in actions. Honors students were often discussed as having the ability to engage in self-monitoring, whereas regular students were often discussed as needing extra support in the form of policing or constant refocus reminders.

Worthy (2010) identified her own racial, ethnic, and SES background

(European American and middle class) as a potential limitation for what she saw and understood when evaluating the data. Answers were coded and three analysis meetings were conducted between the coders, but further strength could have been added to the study if inter-rater reliability between the coders was provided. The level of agreement between coders was unclear.

Sørenson and Hallinan (1986) used a quantitative, longitudinal study of 48 classes of elementary school students, to determine if ability grouping had an effect on academic achievement. A total of 1,477 students took part in the study. Researchers had a sample size of 48 classes: 10 classes were fourth grade, 12 were fifth grade, 10 were sixth grade, five were seventh grade, and 11 were in combined grades. Researchers used racial composition as one factor when selecting these northern California schools. Sørenson and Hallinan (1986) chose reading as their subject of study, because their focus was on academic achievement. Sørenson and Hallinan wrote, "We consider academic achievement in a particular domain, such as reading achievement; and we assume it is measured as a continuous variable, like a score on an achievement test" (p. 523). A group of 148 students (10%) did not take a reading achievement test at the beginning of the school year, but were still included as participants. Data was gathered six times during the year, and teachers discussed the names of students in the collection period, reported out about how resulting groups were formed, and reported out the amount of time that students spent in the ability groups during the collection periods.

Students in 34 of the classes (about 71%) were grouped during reading. Twenty-four of these teachers stated that ability was the criteria used for grouping placement. The Sørenson and Hallinan (1986) study had a mortality rate of 2%, because one class had to be excluded due to a lack of data. A total of 564 students had completed enough data to be analyzed at the end of the school year. Of the 564 students, 384 (68%) were part of classes that were grouped for reading, while 180 (32%) were part of classes that remained ungrouped.

Sørenson and Hallinan (1986) determined that students in grouped classes experienced as much improvement in reading achievement as did students in non-grouped classes. Researchers hypothesized that higher-ability students might have been present in grouped classes. Researchers also hypothesized that there were genuinely positive results from the use of ability grouping. After data collection, researchers used ordinary least-squares regression. Race had a strong impact on achievement, where the estimated effect of race on student ability and effort was .975 in grouped classes and .633 in ungrouped classes. In addition, students in grouped classes were taught less material than students in non-grouped classes

Researchers reported mixed conclusions about the use of ability grouping. First, Sørenson and Hallinan found that students in higher ability groups appeared to have greater opportunities to engage in learning. Second, researchers found that students in grouped classes were taught less than students in ungrouped classes. Third, researchers also reported that students in

small homogeneous ability groups could have learned more than being in larger heterogeneous groups. Based on these conflicting findings, researchers concluded that the overall, the use of ability groups was a detrimental educational practice based on the bias used to group students, the social inequity that resulted from those groups, and the difference in material taught to grouped and ungrouped classes. However, researchers stated, "Our results are dependent on a particular specification of a theory of growth in academic achievement" (Sørenson & Hallinan, 1986, p.541) and speculated that alternative theories about the use of ability grouping could have explained the mixed findings, but the burden of proof was shifted to those who made that argument.

The size of the study indicates that it could be generalized to other elementary schools that had the same or similar model as California schools. Researchers were unsure about how to interpret their results, while the mixed findings may indicate possibilities for future study. The study had some threat to objectivity, in that there was little evidence about how the data was collected, when the data was collected, and the type of data that was collected.

### Alternative Strategies to Tracking Structures

This section will discuss the next steps for proponents of detracking efforts. The term *detracking* refers to efforts to provide for previously excluded groups of students more equal access to different coursework, resources, and potential benefits of track mobility. Detracking usually requires transformation among entire educational institutions and sometimes districts, and success is tied

directly to how much support proponents can receive from a majority of the people involved. Detracking is not easy, but some research indicates that detracking could be worth trying.

Cooperative learning within a heterogeneous classroom could be an effective institutional change. Stevens and Slavin (1995) found that cooperative learning could be sustained when changes were integrated slowly, one at a time. Stevens and Slavin compared cooperative learning schools with traditional schools which used ability grouping or tracking structures. Student tracking based on ability is not the only option. According to Ansalone and Ming (2006), an individualized learning program called Programmed Learning Sequence (PLS) could be beneficial for students who prefer independent work. Researcher noted that the PLS structure could help alleviate pressure on students from competing with peers, as is usually found in tracked classes.

Yonezawa, Wells, and Serna (2002) studied schools that voluntarily used the detracking strategies of decreasing prerequisite classes for more advanced classes. Yonezawa et al. discovered that even when prerequisite classes were decreased, students in lower level classes were unable to gain entry into higher level classes. This study highlighted the amount of effort that is required to make detracking an institutional change that is beneficial to students. Welner and Burris (2006) researched methods to introduce detracking into schools. Welner and Burris found that detracking was successful when the community and the school were on board with proposed changes.

Stevens and Slavin (1995) explored the sustainability of using a cooperative elementary school model over a two-year longitudinal study. The following three research questions were used to frame the study:

- Could cooperative learning be used on a broad scale in many subjects and over extended time periods to fundamentally change the organization of schools and classrooms?

- Would cooperative learning methods stil be effective if they become the primary mode of instruction in schools and would they maintain their effectiveness over time?

- Would schoolwide use of cooperative principles enhance the school's potential to successfully mainstream learning disabled students? (p. 322)

Proposed changes included mainstreaming students identified as being academically handicapped into the cooperative general education classroom, and having students participate in heterogeneous mixed-ability groups. Teachers used cooperative learning in their classes, engaged in peer coaching, collaborated with each other for instructional planning, and collaborated with principals to make decisions about school planning and increasing active parental involvement.

Treatment was administered to schools by phasing in three different

cooperative learning strategies over the two years of the study. Phasing time was from August to March of the first year. Initially, each treatment school received a steering committee, which was a liaison for communication between teachers, researchers, and administrators). Comparison schools were allotted the same amount of instructional time and used the same curricular standards. Faculty in treatment schools participated in development training at the beginning of the first year. Faculty in comparison schools continued to use two or three different ability-based reading groups during classroom instruction.

The Cooperative Integrated Reading and Composition (CIRC) program was introduced in the initial stages of development because it was the most complex system used in the model. Team Assisted Individualization-Mathematics (TAI) was also used.      Programs originated from John Hopkins University's models of cooperative learning. It is important to note that during the first year, only 60% of students identified as being academically disabled were able to be mainstreamed. Comparison schools did not use mainstreaming in their classrooms. In the second year, the decision was made that pull-out remediation was not going to be used for special education. Instead, learning-disabled students received all their instruction in the regular classroom and within heterogeneous groups with their non-disabled peers. In order to achieve this, remedial teachers and other teachers would team-teach to support all the students. Researchers used a sample size of 1,012 students in Grades 2-6 in five elementary schools, in a suburban Maryland school district. Twenty-one

classes from treatment schools were compared with 24 classes from comparison schools, using the California Achievement Test scores in Total Reading, Total Language, and Total Mathematics. Stevens and Slavin (1995) also matched schools based on average student achievement, ethnicity and socioeconomic background.

Pretest scores from the California Achievement standardized test were converted to z-scores. For comparison, teachers administered California Achievement Test (Form E) in subtests of reading, language, math, and reading, during the spring of each year. Reliability for these subtests had a range from 0.80 to 0.95, and the median reliability was 0.91, which indicated a strong level of reliability for the instruments being used. Posttest raw scores were converted to z-scores.

Students were also given attitude measures for pretest/posttest comparison. Pretests were given during the fall of the first year; posttests were given in the spring of the second year. Students were asked to rate their attitudes toward self-perceived abilities in content areas: reading, language arts, and math. Students also rated their interest and ability using a 3-point scale (*like/don't like*, *really good/not very good*). For the pretest attitude measure, the alpha reliability was .35, and for the posttest it was .33. For the pretest perceived ability measure, the alpha reliability was .32, and the posttest was .37.

Researchers used a social relations measure in which students were asked to list the names of friends in the class. Then the average number of

friends listed by the treatment groups was compared with the average number of friends listed by the comparison groups. Test-retest reliability was .57, and for learning-disabled students, the results were analyzed according to the number of times he or she was listed as a friend by a non-disabled peer. The posttest measure was used to determine how and if mainstreamed students had been socially accepted by their peers.

Results included gains in reading vocabulary after the first year (t = 2.14, p < .05), and in reading vocabulary, reading comprehension, language expression, and math computation in the second year for treatment schools. For second-year treatment students, the reading vocabulary t-score was 3.04 (p < .01), reading comprehension t-score was 3.62 (p < .01), language expression t-score was 2.93 (p < .01), and math computation t-score was 3.77 (p < .01). The research method used seems to have higher levels of transferability, due to the fact that students who were identified as being learning disabled were also included in the study. Learning-disabled students in treatment schools listed more friends on the social relations measure than did their counterparts in comparison schools (t = 3.42, p < .01), with an effect size of +.86. Non-disabled students also listed learning-disabled students as being friends more often in treatment schools than in comparison schools (t = 4.33, p < .01), with an effect size of +.90.  More often than not, students identified as academically handicapped are excluded from many studies.

Researchers also included information regarding the academic

performance of gifted students, another oft-neglected group of students. After the second year, gifted students in the treatment schools had higher achievement in reading vocabulary, reading comprehension, language expression, and math computation, when compared with students at comparison schools who participated in pull-out enrichment programs. Gifted students in the cooperative elementary schools also listed having more friends than those in comparison schools (t = 2.64, p < .01). A main argument for the use of tracking or fixed-ability grouping, is that it is beneficial for gifted students. However, this study sheds some light on gifted student performance in mixed-ability classrooms.

Ansalone and Ming (2006) examined the causal relationship between educational institutions and the perpetuation of the achievement gap by the use of student tracking, which limits student capabilities; they studied the academic performance of 71 Bermudan seventh-grade students and found that when the teacher used Programmed Learning Sequence (PLS)—an alternative to student tracking which emphasized self-pacing, active participation, and a variety of auditory and tactile learning strategies—underperforming students had more favorable attitudes toward PLS and scored higher on the PLS instructional assessment than students who received traditional instruction.

Study participants were divided into four groups, and each group received instruction by the same teacher. Researchers assessed learning through comparison of pre- and posttest scores at the beginning and end of four social studies units. Ansalone and Ming (2006) used an instrument called the Semantic

Differential Scale (SDS), which examined the attitudes students held toward each style of teaching and learning: PLS versus traditional. No information was provided about the reliability measure of the scale.

All four seventh grade classes were taught by the same teacher and were given the same social studies content. Researchers wanted to compare PLS to a traditional method of instruction. Each class was given a learning style inventory to determine their learning preferences. All four classes were taught one unit using PLS and one unit using traditional methods over two weeks. At the beginning of each week, students were given a pretest, at the end of each week students were given a posttest. Classes were not distinguished any further than being labeled Class 1, Class 2, Class 3, and Class 4. Students were randomly assigned to each of the four groups. Researchers made no notes about whether or not the school used tracked classes. Researchers introduced PLS to test if it could be a viable alternative to traditional methods of instruction, including teacher lecture and student completion of worksheets (Ansalone & Ming, 2006, p. 7).

During the first week, Class 1 studied the topic of rain forest, from the PLS unit. Class 2 studied the topic of rain forest, from the traditional unit. Class 3 studied the topic of Egypt, from the PLS unit. Class 4 studied the topic of Egypt, from the traditional unit.

During the second week, Class 1 studied Egypt, from the traditional unit. Class 2 studied Egypt, from the PLS unit. Class 3 studied rain forest, from the

traditional unit. Class 4 studied rain forest, from the PLS unit. After the first two weeks, students completed the SDS. After completion of each unit, researchers compared post test scores for the PLS units and the traditional units.

The study was quantitative, and results indicated that underperforming students had more favorable attitudes ($p < .001$) towards PLS over the traditional tracked format of classes. Students were taught first using traditional format, followed by PLS, then traditional, then PLS. Underperforming students also scored significantly higher on the PLS instructional assessment than did underperforming students in traditional instruction ($p < .005$). All students performed higher with PLS than with traditional tracking instruction ($p < .005$), with an effect size of 0.75. The PLS program placed emphasis on instructional strategies including self-pacing, active participation, and the use of other tactile and auditory techniques.

A weakness of this study is that it was conducted in Bermuda, and transferability to other settings (like the United States) may not produce the same findings. However, the approach that was taken in Bermuda, relying on information from learning style inventories and adjusting instruction accordingly, suggests strategies that could be used in United States' classrooms. Researchers concluded that PLS, "is generally not recommended for all children. The PLS is most suited to those who prefer to work alone without the interaction of others" (Ansalone & Ming, 2006, p. 6).

If students have more choice in curriculum, will students choose to move

between different tracks? Yonezawa et al. (2002) examined school  detracking efforts using a qualitative, three-year, longitudinal case study. Researchers gathered data from 10 racially and socioeconomically diverse secondary schools: four middle schools and six high schools. All 10 schools voluntarily adopted detracking strategies to increase the presence of heterogeneous grouping of students based on mixed abilities. Schools used freedom of choice for different academic classes as the detracking strategy; school officials decreased course prerequisites so students could gain access to courses more readily.

Yonezawa et al. (2002) conducted 423 semi-structured interviews that were taped and transcribed verbatim. The study also included 75 classroom observations, where the focus was placed on academic subjects, as opposed to vocational education. Faculty, parent and community meetings were also observed. Researchers collected over 100 artifacts (documents) from the schools, and all information collected was coded and analyzed. Researchers did not include information about inter-rater reliability, how many coders worked with the data, or how answers were coded. The single case-study reports from each school were used in a cross-case analysis, and if a theme was present in the data, researchers re-examined the larger source for potential significance or contradictory evidence.

Classes remained separate, even with decreased prerequisites, due to reported hidden institutional barriers. Yonezawa et al. (2002) found three different ways that social status had an effect on course placement. First,

different groups received uneven information based on the form and content of the information network. Second, schools seemed to exhibit selective flexibility, where schools would alter their course offerings to match different characteristics of their students. Third, students were blocked from picking higher-level courses due to hidden prerequisites. For example, if students wanted to qualify to take one course, they needed to demonstrate knowledge of information that was presented in previous courses, where schools largely downplayed that information when presenting detracking strategies.

Yonezawa et al. (2002) found that the use of detracking helped solidify lower-track students' self-concept about being unable to compete with higher-track students, and helped solidify higher-track students' self-view of entitlement or privilege. More importantly, students chose to remain in lower-track classes, and they desired to learn in environments that were respectful. Researchers' findings about the desire to learn in safe spaces corresponded with those of previous studies. The study had credibility in that the findings are believable and convincing, as researchers were transparent about data-collection and demographic diversity of school participants. However, Yonezawa et al. (2002) did not mention checking with members of the study before publishing study results, did not mention how answers were coded and if inter-rater reliability existed between coders, all of which decreased credibility of the study. The study had high transferability given the demographic diversity present. There was high confirmability, as an outside party would have been able to audit the process and

product of the study.

When schools attempt to detrack, political controversy is always present. Welner and Burris (2006) explored two different approaches that had been used to tackle the politics surrounding detracking. Researchers focused on answering the question: How can we use existing data to defend the importance of school detracking against arguments of potential negative impact on others? Two case studies were used as exemplars of the detracking approaches. The first was in Long Island, New York, and represented a gradual integration of change strategy. A co-author of this first case study was a principal at the high school during a critical detracking point in the changeover. The second case study focused on a school in Pittsburgh, Pennsylvania, and represented a rapid integration of change strategy.

In the first case study, changes were phased in slowly. During the first phase, school leaders realized that the school-level track (the low track) was not providing an effective education for all students, and tracks were divided according to race and SES. Three distinct levels existed: 1) school-level, 2) Regents, and 3) honors. Students took Regents courses to qualify for the New York State Regents diploma. Parents were given the option of choosing a higher level Regents course for their students, to help diminish the stratification effects, and all students were eligible for enrollment in the Regents courses. Regents courses became more heterogeneous. In the first phase of detracking, teachers also worked cooperatively to create rubrics that evaluated students on both

process and product.

The second phase occurred in 1998. School officials had data to support the implementation of phase two, which included data from both high school and middle school detracking efforts. Ninth-grade English and Social Studies classes were detracked during 1998-1999, and all students were placed in heterogeneous classes. As principal of the school, co-author Burris, also introduced a heterogeneous track in ninth-grade Math and Science. School leaders were presented with the detracking data and felt it should continue to include the 10$^{th}$-grade curriculum in both English and Social Studies. Parents pushed back and resisted against the proposed change. Teachers and school leaders discussed with parents how the new classes would be structured, and leaders presented parents with data about academic achievement. Student scores for the New York State 10$^{th}$-grade Regents Exam increased at the end of the school year, when the 10$^{th}$-grade had been detracked.

In 2000, after the first phase had concluded, 82% of graduates had earned a New York State Regents diploma. Welner and Burris (2006)  found ninth-grade class culture had visibly changed after the conclusion of phase two (detracking the ninth-grade curriculum). Minority students had higher expectations and performance levels than in previous years. During the second phase, the curriculum did not lower the cognitive demand of tasks assigned to students, and academic achievement by those who were identified as being high achievers did not decrease.

Parties involved in the second case study took a vastly different approach. Community members and parents in Woodland Hills, Pennsylvania, sought court-mandated detracking, after it was discovered that tracking in the district was creating racial segregation within legally desegregated schools. During the 1996-1997 school year, 6000 students were enrolled in the nine schools that made up the district; 27% were African American, and 70% were European American.

Since there was a small group individuals with vested interests, they had to pursue legal action because they lacked the critical mass that was present in the New York case study. Federal district court found that tracking was discriminatory and ordered that it be eliminated. As a result, the superintendent detracked English courses from 1995-1996. To appease other concerned parents, higher-track classes were added back in, but the courses were opened up to students of color, although higher-achieving students were pushed upward into higher courses. Findings from this second case study included that students were able to move to more educative environments. The open enrollment for classes also helped decrease the frequency of occurrence of students backsliding into lower goals and lower expectations.

Researchers noted, "The two approaches described here are not mutually exclusive... the two approaches do differ markedly in terms of the steps needed to initiate and sustain the reform effort" (Welner & Burris, 2006, p. 98). Researchers concluded that the first gradual introduction of change with school support was favorable because it helped shift trust back to educators, especially

during unknown times of change. Researchers concluded that the second rapid introduction of change with community support was necessary when eduction is seen as a community asset. Burris' direct participation in the first case detracking strategy must be addressed. While the researchers appeared to objectively evaluate each strategy, a hierarchy of importance seemed to be present: the first case study detracking strategy was favorable to the second case study detracking strategy. In addition, the first strategy had detailed notes outlining the strategy, while the second strategy had scarce notes outlining the strategy. While student data was provided from the first case study, no student data was provided from the second case study.

A strength of the two case studies include a discussion about how equity pedagogy was used as a foundation to build commitment for the cause of educational attainment. A weakness of the case study was that no data was provided, so there was a lack of credibility, dependability, transferability, and confirmability.

## Summary

Student tracking is a widely debated institutional practice. This chapter explored research from topics at the individual level (teacher and student) and at the institutional level (schools). For teachers, perceptions and different instructional methods based on student grouping are factors that need to be taken into account when making decisions about instructional practice. For students, perceptions and comparisons are factors that need to be taken into

112

account if tracking structures exist. Other factors for consideration are outside factors, positive results from student tracking, negative results from student tracking, mixed results from student tracking, and detracking strategies used at the institutional level. All of the categories under which research was reviewed demonstrate the complexity of the student tracking and why so much debate surrounds the use of tracking within public schools.

CHAPTER 3: CONCLUSION

**Introduction**

What are possible negative effects that student tracking could have on elementary age students within the United States? Institutional practices, including the use of a hidden curriculum that limits access to families and students who do not readily know codes of power, tend to allow students to be categorized as low-achieving. Low-achieving students run the risk of being placed in lower-track classes. Tracking refers to rigid groupings that often remove students from heterogeneous classrooms and place them within homogeneous classrooms with peers of similar abilities.

Track placement is determined by perceived levels of academic ability or intelligence. A dangerous implication is that these perceived levels are fixed for each student, as indicated by a lack of mobility between tracks. Potentially harmful assumptions lie with association with different tracks.  Examples of such assumptions are that students in lower tracks have lower motivations, that students in lower tracks do not have the ability to succeed in higher education, and that students in lower tracks should concentrate on vocations. Social mobility is also limited; students have the tendency to compare themselves to, and socialize exclusively with, students from their track or level.

Student tracking is often seen most clearly during high school and predominantly in subjects like math or science. However, at the elementary level,

students are often being grouped based on ability, which could seal their academic fate in subsequent years. In Chapter 1, this paper explored the rise of student tracking. By the 1920s, the population had changed, and more students were looking toward vocations rather than higher education. As a result, the schools changed, developing tracks that would help students reach their career or educational goals. School funding became tied to vocational offerings, and teachers were required to have studied the vocation of the classes they were going to teach. Over time, the practice of student tracking changed from being beneficial to students, to something covert and socially divisive. Students placed in lower tracks are often minorities, and teachers in lower tracks often rely on instructional practices that are used to catch students up, such as product questions or memorization questions.

The contemporary term *ability grouping* is used as a replacement for the term *tracking*; however, students are still being grouped according to perceived ability. At the elementary level this can be especially harmful, as teacher perception plays a much larger role than at subsequent grade levels. In Chapter 2, this paper explored how different factors—social comparisons, teacher perceived ability, student perceptions, and teacher perceptions—all contribute to the complexity around the use of student tracking. Since most tracking structures do not begin until late middle school or early high school, teachers rely on factors such as student behavior, student socialization skills, and student aptitude—largely based on standardized test results—to determine where a

student falls on an attainment scale. In Chapter 3, this paper will give an overall summary of the findings, implications for classroom practice, and suggestions for future research.

## Summary of Findings

This section reviews the major findings of the thirty peer-reviewed research studies covered in this paper. The majority of the qualitative research included in this paper was conducted using a case-study format. The majority of quantitative research was correlation analysis, and ANOVA or MANCOVA was used for statistical analysis. The following summary of findings includes major themes across the different findings. Overall, analysis of the research has revealed that perception plays a large role in teacher and student perceived levels of ability. Students are less likely to compare themselves with students from different tracks, and students can change their level of self-concept based on track placement. Teachers are less likely to assign periods of independent choice to perceived lower tracks, and are more likely to rely on instructional practices, such as direct instruction, that lower the cognitive demand of subjects in perceived lower-track classes.  Female European-American teachers were seen throughout a majority of these studies. Students in fourth, fifth, sixth, and seventh grades were the research focus groups for many of the elementary and middle school studies. With the exception of identified urban elementary schools or urban areas, the majority of research studies had an oversampling of European-American students. Finally, detracking can be difficult, but can be

successful if enough support and planning is built into the strategy; entire institutions can be changed to increase equity across different student populations.

## Teacher Perceptions

Leder (1987), Parsons et al. (1982), Neumeister et al. (2007), Haller and Davis (1981), and Monsaas and Engelhard (1994) conducted studies that examined the influence of teacher perceptions in the classroom. Teacher perceptions of ability are largely influential in track placement decisions. Leder (1987) determined that the type of interaction varied from achievement group to achievement group. The teacher also changed the type of question asked from group to group, indicative of an inherent bias about ability from one achievement group to the next. While the findings seem to be aligned with this paper's research question about the negative effects of student tracking on elementary aged students, Leder's (1987) case study cannot stand alone based on the lack of participant demographic information and small sample size, which decreased the overall credibility of the study.

However, quantitative analysis of questionnaire data by Parsons et al. (1982) could have laid the foundation for Leder's (1987) study. Researchers were interested if student-teacher interactions changed  across achievement groups. Additionally, researchers wanted to investigate if the changes had an impact on student self-concept related to ability. Parsons et al. found that student-teacher interactions changed based on gender. While male students received the most

widely varied treatment, female students reported the highest impact based on varied treatment. Results are indicative of the powerful impact interactions have on the classroom learning environment. Parsons et al. had high reliability based on the transparency of coding of the data, the number of independent coders, and the level of inter-rater reliability between coders. Parsons et al. (1982) and Leder (1987) both demonstrated how teacher-student interactions can vary from one ability group to another.

Haller and Davis (1981) discovered "when teacher judgment is used to form groups, the degree of socioeconomic segregation, on the average, increased about 20 percent" (p. 167). Researchers inferred that student SES could have been a stronger factor than student reading ability was for grouping decisions. Neumeister et al. (2007) discovered that teachers of predominantly minority student populations often had difficulty identifying gifted characteristics that were not from the dominant social culture. Behaviors such as being artistic, showing boredom, or showing disinterest, were not readily mentioned by 23 of the 27 (about 85%) participating elementary teachers. The study was critical because it addressed the criticism that there is an oversampling of minority students in lower-track classes. Haller and Davis (1981) found teachers did not mention reading skill as an indicator for group placement about 25% of the time, which is indicative that teachers were relying on other factors for their decisions. This assertion was partly confirmed by the results found by Neumeister et al. (2007), which indicated that teacher perceptions about identification of student

giftedness were aligned closely with the dominant social culture, including socially accepted behaviors of identified gifted students. Neumeister et al. (2007) had high credibility that the findings were believable and convincing based on the research method. However, confirmability could have been increased by including information about inter-rater reliability.

Monsaas and Engelhard (1994) examined test preparation strategies and found that elementary teachers engaged in higher levels of test preparation behaviors than teachers at other grade levels. Elementary teachers seemed to have the perception of a higher vested interested in academic success, because they were single teachers responsible for a classroom of students). Monsaas and Engelhard yielded interesting teacher perception results, although a weakness existed in that all teachers who participated in the study were pursuing advanced educational degrees.

One pattern was that Haller and Davis (1981), Neumeister et al. (2007), and Monsaas and Engelhard (1994) all had high levels of racial homogeneity; Haller and Davis (1981) had it in the oversampling of European American students, while Neumeister et al. (2007) and Monsaas and Engelhard (1994) had it in the oversampling of Caucasian teachers. According to Washington state's Office of Superintendent of Public Instruction, in the 2009-10 school year, 91.89% of elementary school teachers in the state of Washington were White (Office of Superintendent of Public Instruction, p. 1). Reported numbers mirrored the statistics found in the sample by Neumeister et al. (2007) and Monsaas and

Engelhard (1994). The entire group of studies under this heading have demonstrated the way in which teacher perception is influential to instructional decision-making.

Instructional methods change based on student grouping. Waxman et al. (1997) reported that students in efficient schools spent 72% of instructional time in a larger group setting, while students in inefficient schools spent 81% of instructional time in a larger group setting. Researchers also noted that in efficient schools, students spent 40% of time watching or listening, and in inefficient schools, students spent 52% of time watching or listening. These results could indicate a higher level of teacher reliance on direct instruction in inefficient schools; these students spent more time listening than engaging in active learning. Waxman et al. also demonstrated that tracking can occur between schools, where E/E schools could represent higher track schools and I/I schools could represent lower track schools.

Eder (1981) studied teacher-student interactions, but focused on grouping within one specific classroom, within one specific subject: reading. While Eder found that the teacher had more positive interactions with students in the higher-ability groups than with those in lower-ability groups, the case study focused on only one classroom, within one school, and lacked any demographic information. Kersten (2007) also studied a small group of students, in a similar longitudinal case study. Kersten found that the gifted student reported a variety of interesting subject matter, while her grade-level classmates reported a lack of interesting

subject matter; grade-level students reported that Silent Sustained Reading (SSR) was their favorite time, because they were able to choose their reading materials. Kersten (2007) had higher credibility during her one-year case study, because she included a discussion about data-collection and research methodology.

As a whole, the research studies in this section demonstrated the way in which instructional practices differ across achievement groups. Waxman et al. (1997) determined school efficiency partly based on achievement test scores, Kersten (2007) determined that student choice in subject matter changed across different ability groups, and Eder (1981) determined that teachers had more positive interactions with students in higher ability groups. Taken together, these studies indicate that there are differences in instructional practices based on ability-group placements.

<center>**Student Perceptions**</center>

Crumpton and Gregory (2011) found that students who were engaged with the material and found it relevant to their lives also reported having higher levels of overall motivation. Strength was added to the study because of the transparency of the methodology, which increased study objectivity. Butler (2008) reported that students in both high-track and low-track classes were unwilling to seek help during math class, because of the detrimental effect it would have on social comparisons. Similarly, Krebs and Roebers (2010) discovered that students were moderately aware when they tried to identify correct versus

incorrect answers. Hulan (2010) observed that students changed their interactions with each other and the material studied, when the teacher was present and trying to dominate the small-group discussion. Hulan also found that students and their teacher differed in perceptions about the material, where the teacher was convinced the students loved the curricular materials, while the students did not mention them at all. Brint et al. (2001) found that students who were able to make inferences about social expectations had more success in primary schools than did students who were less able to make inferences about social expectations. Skybo and Buck (2007) observed that students reported higher levels of stress during times of standardized testing, with symptoms including hunger and headaches.

Collectively, the studies demonstrate that student perceptions about testing, group placement, and academic relevancy were strong influences on overall student behavior.

Students can perceive comparisons between themselves and other students. Chiu et al. (2008) examined student self-concept, or how much students felt they knew about the content area, and found that student track placement largely affected math self-concept. As a result, students were more likely to make within-class social comparisons than between-class social comparisons. Dumas et al. (2005) also explored social comparison at the elementary level. Researchers determined that students were less likely to compare themselves with students who were academically outperforming them in

different courses. Bain and Bell (2004) looked at gifted and high-achieving students and noted that gifted students reported higher levels of social self-concept than high-achieving students. Bain and Bell also reported that gifted students connected high self-concept with high ability more often than did high achieving students. The three studies in this category all explored student social comparisons. An emerging pattern was that students did not readily make comparisons across tracks.

## Outside Factors

Patrick et al. (2005) found that parental monitoring during early elementary school years was negatively correlated with covert and overt conduct problems in later elementary school years. Student conduct is one factor that is sometimes used to determine track placement in later academic years. Patrick et al. (2005) added collective strength to their longitudinal study by including transparency about coding and reliability measures used.

DeSena and Ansalone (2009) determined that gentrification had changed the perceived quality of neighborhood schools, and parents showed preference toward charter schools over neighborhood schools. Thompson et al. (2011) explored student mobility and found that there was statistical significance between school poverty and student mobility; reading was the most negatively correlated variable with student mobility. For example, when student mobility was lower, school reading scores were higher. Researchers used a large sample size, which was a strength of the study because it helped account for variance

within the sample population. While DeSena and Ansalone (2009) gave valuable insight into gentrification in the New York state area, the lack of information about data-collection and small sample size decreased the overall credibility of the study. DeSena and Ansalone noted that parents claimed that neighborhood schools taught to the tests, and opted to send students to alternative schools. Interestingly, Lopata et al. (2005) compared Montessori programs to other traditional educational programs and came up with mixed findings about researchers' hypothesized correlation between higher academic achievement and enrollment in Montessori programs. Montessori programs were based on the idea that students should experience large periods of individually chosen activity and should not be subjected to specific pacing; rather, students should largely determine their own individual pace. Findings indicate that decreased formal structure may not necessarily be conducive to academic achievement.

## Positive Results from Tracking Structures

McCoach et al. (2006) determined that kindergarten students who were placed in within-class ability groups had reading gains during their school year. An inference based on the findings is that the use of homogeneous ability grouping may not be harmful in and of itself. At the earliest grade level, students gained in reading from the educational practice in their classroom. While this finding conflicts with this paper's research question about the negative effects student tracking has on elementary aged students, it gives valuable data about how homogeneous grouping may be beneficial for students. The questions

124

remain of how and when this strategy can be used successfully. Further, Lyle (1999) also found that mixed-ability grouping was beneficial for all students. Students from different attainment groups reported increased group work skills and increased understanding of that material being worked with. Even though Lyle (1999) studied mixed-ability grouping, the initial assignment of the groups was made based on comparison of ability between students. Lyle (1999) and McCoach et al. (2006) demonstrated the importance of mindful grouping strategies when using ability grouping.

## Negative Results from Tracking Structures

What do most people think about when they hear the term *tracking*? Worthy (2010) observed that teachers varied instructional methods greatly between honors and grade-level courses. While students in honors courses were engaged with creative, analytical assignments, students in grade-level courses had to work exclusively with basal readers and worksheets. LeTendre et al. (2003) explored how tracking was used in the United States, Germany, and Japan, and found that students were tracked based on perceived ability in all three countries. Researchers also found negative impacts on minority students, immigrants, and lower-SES families, that were comparable across all three countries. Worthy (2010) and LeTendre et al. (2003) demonstrated findings that have been touted by detracking proponents: students across different tracks have different access to high-quality education, and the inequity could significantly impact students in the future.

125

Student tracking is complex, which is evidenced by the next study. Sørenson and Hallinan (1986) discovered students in grouped classes had the same level of reading improvement as students in non-grouped classes. Research findings are indicative of why student tracking is such a widely debated topic; many mixed and/or conflicting findings exist. In the case of Sørenson and Hallinan (1986), the internal validity of the longitudinal case study was decreased, because researchers could not conclude whether or not the variable of ability grouping caused a change in the variable of student reading performance.

## Alternative Strategies to Tracking Structures

How do schools go about detracking? Ansalone and Ming (2006), Stevens and Slavin (1995), Yonezawa et al. (2002), and Welner and Burris (2006) sought to discover alternatives to traditional tracking structures. Stevens and Slavin (1995) conducted their research in the sustainability of a cooperative elementary school, and found that students across different populations experienced social and academic gains. Researchers discovered that students who were considered to be learning disabled—a population that was often neglected by previous detracking research—were successfully mainstreamed into the cooperative learning elementary school. Stevens and Slavin (1995) reported the school-wide support needed to initiate and sustain institutional change. Welner and Burris (2006) discussed similar results in the form of two detracking strategies a gradual implementation of change and a rapid implementation of change. In both case

studies, schools reported overall better learning communities.

Ansalone and Ming (2006) examined an alternative instructional program that allowed for student self-pacing and increased student engagement through different tactile and auditory strategies. Researchers found that students favored the alternative Programmed Learning Sequence (PLS), and under-performing students scored higher on the PLS assessments than on traditional assessments. While the research was conducted with a small sample size in Bermuda, the study could be replicated in the United States to explore alternative methods for instruction and assessment. Yonezawa et al. (2002) demonstrated how detracking efforts needed to be well planned and transparent. Researchers found when school officials removed prerequisites for higher-track classes, students did not choose to move to those classes. Yonezawa et al. (2002) found that parents and students felt that detracking efforts were undermined by hidden prerequisites, which had not been factored into during the implementation of a detracking plan. While Yonezawa et al. (2002) and Welner and Burris (2006) conducted their research at the high school level; the results had strong implications for educational leaders at the elementary level: plan well, be transparent, and gain as much support as possible (which could include using data-driven research).

**Classroom Implications**

Institutional change requires time and additional support. Stevens and Slavin (1995) demonstrated the importance of allowing for additional time and

127

support when implementing school-wide changes. Researchers waited for two months between phasing in new components of the model. Teachers were given feedback about ways to enhance their instruction, and they were actively involved in collaboration across content areas and along hierarchical school structures, in conjunction with school principal and administrators. For teachers, allowing time for changes to be implemented first before providing new training may have been one of the reasons the structural changes were successful. Teachers and students were given the chance to work with the new structures for a while before adding new components. Teachers are often frontloaded with new information to inform instructional practices at the beginning of the year, and are asked to build in changes without additional feedback or support. The study sheds light on how important communication is between all parties involved, if changes in structural, curricular, and instructional practices are going to be implemented. While educational reform can be motivated by the best of intentions, teachers, support staff, and administrators all need the necessary resources—including additional time—to make true institutional change feasible.

Another interesting finding by Stevens and Slavin (1995) was that gifted students experienced academic improvement and reported more friendships in cooperative elementary schools than in comparison schools. Team-teaching with special education teachers, collaboration with principals, and opportunities for increased parental involvement are components that contributed to the success of that cooperative elementary school model. The use of those educational

practices could be beneficial across all student populations. This study was particularly relevant because the other studies did not include data about students who were considered learning disabled. Cross-functional teams of educators are necessary to make institutional reform possible. Decreased delineations around different groups of teachers, administrators, and staff could help increase overall morale for the learning community.

Hulan (2010) discovered that a teacher and her students had different perspectives about how the use of prescriptive reading texts was received by readers. The teacher asserted that students "like the Rigby books," while none of the students self-reported an interest in greater use of the texts (Hulan, 2010, p. 60). The teacher also stated that students preferred fiction to nonfiction, while Hulan had field notes that documented high levels of student engagement with nonfiction texts concerning the human body and animals. This study demonstrated the importance of starting where students are, rather than where teachers want or expect them to be. It is clear that the teacher thought that the students generally cared for the texts, due to an absence of objection or negative statements about the texts. A classroom practice that can be extended from the study is developing a routine about how to discuss texts. Students will not automatically know how and when to discuss texts; teachers can set guidelines for respectful discussion, which can be beneficial to both guided groups and the larger class as ways to build community classroom.

Waxman et al. (1997) found that students in Ineffective/Inefficient (I/I)

schools spent less time on writing assignments than did Effective/Efficient (E/E) comparison school students. Another relevant finding was that students in I/I schools spent more time in passively listening in large-group settings than did students in E/E schools. Teachers can use smaller learning groups to help increase student engagement and to gain a better control over pacing for curriculum materials. While whole group instruction has its merits, students stand to benefit from smaller group instruction, where they have a greater chance to actively interact with the subject matter.

Crumpton and Gregory (2011) found that when students reported higher levels of academic relevance, they also reported higher levels of intrinsic motivation. While researchers noted that these two factors alone could not be considered predictive of future academic success, it is important to recognize how academic relevance and motivation could lead to higher self-concept for students. Crumpton and Gregory also found evidence that was contradictory to the perception that students who were deemed low-achievers were essentially unmotivated.

Leder (1987) found that students who were deemed low-achieving, were often asked questions but given little wait time to formulate answers. Additionally, the teacher would often move from lower-achieving students to higher-achieving students after a few incorrect answers were given, in order to gain the correct answer. Educational practices, including decreased wait time, can limit student engagement with the material. Teachers run the risk of alienating students from

both the classroom and the subject matter. Teachers could use problem-based instruction or group-worthy tasks to enhance academic self-concept for all students, including those designated as high- and low-achieving students.

Teachers can ground abstract concepts in relevant examples, drawing from the immediate context found in their classrooms. If teachers used examples from plausible scenarios in students' lives, stronger academic relevancy could be established. Crumpton and Gregory (2011) wrote, "Schools with students who are low-achieving may consider developing ways to make classroom curricula more related to what students find important" (p. 50).

Ansalone and Ming (1996) found that an alternative instructional format was favored by underperforming students. Researchers noted that Programmed Learning Sequence (PLS) was not intended to be used for all students in a classroom. "The PLS is most suited to those who prefer to work alone without the interaction of others. It is ideally suited for the persistent and motivated learner—the one who might use the material until the program has been completed" (Ansalone & Ming, 1996, p. 6). The passage indicates that PLS, which relied on self-pacing and was based on different learning styles, could be worth exploring as an alternative to between-class ability grouping.

Brint et al., (2007) discovered a disparity in multicultural education for students in second and fifth grades. In second grade, curriculum materials displayed messages of cultural acceptance, multiculturalism, and positive aspects of diversity. In fifth-grade, curriculum materials (especially in social

studies), the themes were similar but were subjugated to themes like technological advances and social customs. LeTendre et al (2003) discovered that across the three countries of Germany, Japan, and U.S., students of minority status were placed in lower tracks, and placement was related to student language acquisition rate. Taken together, these two studies could indicate that one negative effect of the use of tracking high status is given to students in higher tracked classes, which extends to the classroom environment itself.

The students in Lyle's (1999) study felt their task had a shared, important purpose and that they were discussing something that was relevant, as opposed to doing assignments that would just be turned in. Students come to school with a myriad of experiences; how much value do schools place on these experiences, especially when these experiences come from cultures that are not considered to be part of the dominant culture? Students possess various communication styles, exhibit different behaviors, and have different ideas about what schooling is, all of which are influenced by the environment in which the student is living or was raised. Teachers need to push academic relevance and be mindful of opportunities to include all students by creating examples or situations where all students are positioned as competent. Students may compare less and focus more when they have a sense of personal agency. Personal agency increases when a student has the motivation to contribute something, or believes that personal effort is tied directly with progress toward a goal. Teachers can help create opportunities for students to establish personal

agency through instructional practices like inquiry-driven learning, or problem-based instruction. Personal agency is especially important when teachers are considering differentiating instruction: How can each activity lend itself toward the goal of students becoming more independent and having more personal agency?

## Suggestions for Further Research

Butler (2008) suggested that teachers may adjust their educational practices in relation to their own self-concept in subject areas. Butler discussed her 2007 study, which determined that when teachers were engaging in ego-avoidance (related to their teaching abilities), students "were more likely to report that the teacher inhibited student questioning and help-seeking by conveying that these behaviors are signs of inadequate ability" (p. 20). Teachers must become reflective practitioners, and further investigation about compensation strategies used by lower self-concept and higher self-concept instructors would be beneficial to the educational community. Critics of the use of tracking have speculated that lower-track classes are subjected to more rote memorization, or to questions from the lowest level of Bloom's taxonomy (knowledge). Researchers could ask the question, "Where are these curriculum choices originating? Are they coming from the needs of the students or the needs of the teachers?"

Further research could be conducted using Butler's (2008) methodology; however, a modification might be needed if researchers are planning to compare schools that have tracking structures across multiple subjects, which could

change the results. Since Butler's (2008) study was conducted in Israel, adaptations would need to be made for study in the United States. For example, Israeli schools used teacher recommendation when placing students, rather than relying students' standardized test scores. In order to replicate the study, researchers would have to collect qualitative data based on teacher perceptions for group placement, similar to that gathered by Haller and Davis (1981) with the card-sorting activity paired with informal interview. Then, correlation could be calculated among variables after similar qualitative scales were administered to students.

Chiu et al. (2008) found that the use of grades was a stronger influence on student self-concept than placement in different math tracks. Further exploration about the use of grades, alternative methods of evaluation, and student self-concept would be beneficial to the academic community because it could shed light on motivating factors for students when completing academic work. Chiu et al. (2008) investigated self-concept at the seventh-grade age level. Conducting similar studies at the early elementary levels would be beneficial to the academic community, because in early elementary grades, students have had little experience with tracking structures. Therefore, comparative correlation analysis could be performed between self-concepts held by student populations with little experience with tracking and self-concepts held by student populations with more experience with tracking. Such studies would help the academic community see how time spent in tracked classes can affect student self-concept. Educational

research is driven by the desire to find best educational practices, and it is important to investigate if self-concept could change over time as a result of tracking.

McCoach et al. (2006), discovered that the use of ability grouping could have been responsible for reading gains in kindergarten students. This study helped demonstrate the need for further research about the use of homogeneous student grouping at the elementary level. Since the use of tracking does not become readily apparent until secondary grade levels, the predicative variables are often harder to determine at the elementary level. To build on McCoach et al., researchers could investigate the types of variables that are consistent across elementary grade level students and secondary level tracked students. Which factors are teachers and administrators using as predictors for future academic achievement? Is there a correlation between the factors listed at the elementary level and the factors stated by teachers at the secondary level?

Ansalone and Ming (2006) discussed the use of PLS as an alternative to student tracking. The researchers noted that this type of instruction should not be used for all students. Further research could be done concerning students with special needs and learning disabilities using a PLS-type of program, which relied on self-pacing and definitions and pictures related to vocabulary. Since Ansalone and Ming conducted their research with Bermudan seventh graders, it would be worthwhile to conduct similar research in the United States. Researchers were interested in strategies to help close the achievement gap, which is a shared goal

in the United States. In order to conduct a similar quantitative, correlation study, researchers would need to find classes that were receiving instruction from the same teacher, using both the alternative program and traditional instructional methods. An adaptation that could be made would be to include team-teachers who had similar instructional methods, and were willing to teach both an alternative and traditional class in the same subject. A similar study would require extra planning, but could yield valuable insight into programs teachers could use that could help them differentiate instruction.

Brint et al. (2007) found that, more often than not, teachers relied on indirect norming expectations in elementary schools. This delivery of socialization messages could limit access to behavioral expectations for students who come from cultural or social groups that do not place emphasis on, or recognize the passive construction of, these messages. Students could be disadvantaged if they only recognize direct messages as norming behavioral expectations. Learning-disabled students or students who struggle with abstraction may also be denied access to necessary socialization codes because of convoluted messages. Researchers noted that "many teachers told us that criticism and praise were not necessary because 'students know the rules'" (Brint et al., 2007, p. 163). Two essential follow-up questions could have be explored: Which students know the rules? In what ways do they display their understanding of the implicit rules?

Worthy (2010) concluded that contemporary ability grouping is aligned

with the definition used for tracking students within the 1970s and 1980s. Worthy's findings and assertions demonstrate how important placement decisions are if ability grouping or tracking is being used. Teachers admitted to being driven by subjective perceptions when making decisions about student ability. The study could be a jumping-off point for future research on ability grouping, including investigation into the question, "Does teacher self-concept affect the choice of instructional methods?" Worthy (2010) also identified the need for parents to be included in discussions about ability grouping and differentiated instructional methods. Yonezawa et al. (2002) confirmed the need for transparency in decision-making; participant feedback included comments about hidden barriers, including lack of access to information and covert prerequisites to gain entry into honors (higher-track) courses.

Skybo and Buck (2007) raised important questions about the type of stress that occurs from the use of proficiency testing. Researchers attempted to answer the question of how stress manifests itself in fourth-grade elementary students. If studied further, this topic could provide helpful information about behaviors that could be indicative of high levels of stress around testing. All of these behaviors may look or sound different depending on the individual student.

**Conclusion**

Student tracking has changed since its introduction in the 1920s. Alternative, vocational track offerings were integrated into schools to meet the changing demographics of the schools and workforce. Federal funding was tied

to vocational class offerings, so that students who were not pursuing higher education had knowledge in subjects that would further their careers. Student tracking slowly changed into a system that separated students based on access to equitable education. Student track placement dictated whether or not students had access to higher education. Tracking became more covert, and the use of variables such as student behavior and perceived ability became downplayed. Tracking then was called something new—ability grouping—but it still relied on dividing students into groups based on perceived ability.

Opponents of the use of student tracking argue that there are disproportionate populations of low-SES and minority students who are forced to settle for less than adequate education based on lower perceived ability. The use of standardized testing also can be demographically isolating for many student populations. Proponents of the use of student tracking assert that students learn best when they are grouped with peers of similar abilities. Proponents also emphasize that legislation such as No Child Left Behind (2001) focused on bringing all students up to standard, and higher-achieving or gifted students were not considered during the decision-making process. Research reviewed in Chapter 2 demonstrated that teachers often relied heavily on perceptions when making decisions about students. Teachers often determined perceived ability by using subjective reasons, such as social interactions, academic performance in content areas, SES, disability status, gender, and overt behaviors. Teachers often would group based on perceived ability in subjects such as reading and

math. Teachers also relied on higher levels of direct instruction, more product-based questions, and more monitoring strategies for students in lower-track classes or lower-ability groups, than in higher-track classes.

Students often had different levels of self-concept and self-esteem which were related to track placement. Students also held strong perceptions about teacher interaction, and about the use of both across-track and within-track comparisons with peers. Students were more likely to engage in within-track comparisons than in across-track comparisons. Student self-concept ranged from entitlement or privilege (higher track), to inability to compete or deficient (lower track). Outside factors included environmental conditions like parental warmth, communication, and monitoring strategies used at home. These factors helped paint a richer picture of the multitude of influential forces students may face and bring with them into the classroom.

Complicating matters further, one study found that ability grouping during reading instruction allowed for student gains across all ability levels. However, ability grouping deterred student growth across ability levels in another study. Researchers were unclear about how best to interpret findings in yet another study. The mixed, sometimes conflicting, research findings about the use of tracking and ability grouping demonstrate why these instructional practices should be researched further.

Alternatives to tracking or ability group structures include heterogeneous, flexible, mixed-ability grouping (with students of different abilities grouped

together), and institution-wide strategies to implement changes for more equitable access to education.

Why are practices used that track students based on perceived academic achievement or ability? Who really benefits from the use of student tracking? These guiding questions are large in scope, broad in nature, and evoke emotive responses from all parties involved. As reflective practitioners, teachers must be able to ask, "Who are instructional methods serving? Who are instructional methods excluding?" Teachers also fall under scrutiny when they work within institutions that use student tracking or are in the process of detracking. Social status is directly related to track placement, and track placement has shifted from its original alternative to higher education, to something that decreases access to quality education for some groups of students. Detracking efforts can be chaotic and tumultuous, and they require buy-in from multiple levels of vested stakeholders. However, recent detracking research has demonstrated that, with careful, strategic planning, and transparency, detracking strategies can be beneficial for diverse student populations and lead to increased academic achievement and more team-oriented learning environments.

REFERENCES

Ability grouping. (2002, July). *The Balanced View*. Retrieved from
htt://www.sharing success.org/code/bv/abilitygrouping.pdf

American President: Theodore Roosevelt. (n.d.) In *American president: A reference resource*. Retrieved from
http://millercenter.org/president/roosevelt/essays/ biography/4

Ansalone, G., & Ming, C. (2006). Programming students for academic success: The PLS—an alternative to traditional tracking. *Educational Research Quarterly*, *29*(3), 3-10.

Bain, S., & Bell, S. (2004). Social self-concept, social attributions, and peer relationships in fourth, fifth, and six graders who are gifted compared to high achievers. *The Gifted Child Quarterly*, *48*(3), 167-178.

Brint, S., Contreras, M., & Matthews, M. (2001). Socialization messages in primary schools: An organizational analysis. *Sociology of Education, 74*(3), 157-80.

Bureau of Education, Department of the Interior (1918). *Cardinal principles of secondary education*. Retrieved from
http://www.archive.org/details/cardinalprincipl00natiuoft

Butler, R. (2008). Ego-involving and frame of reference effects of tracking on elementary school students' motivational orientations and help seeking in math class. *Social Psychology Education*, *11*(1), 5-23.

Chiu, D., Beru, Y., Watley, E., Wubu, S., Simson, E., Kessinger, R., Rivera, A., Schmidlein, P., & Wigfield, A. (2008). Influences of math tracking on seventh-grade students' self-beliefs and social comparisons. *Journal of Educational  Research*, *102*(2), 125-136.

Cornbleth, Catherine. *School curriculum—Hidden curriculum*. (n.d.) Retrieved from < a href="http://education.stateuniversity.com/pages/1899/Curriculum School-HIDDEN-CURRICULUM.html">School Curriculum – Hidden Curriculum < /a>

Crumpton, H., & Gregory A. (2011), "'I'm not learning': The role of academic relevancy for low-achieving students. *Journal of Educational Research, 104*(1), 42-53.

DeSena, J., & Ansalone, G. (2009). Gentrification, schooling and social inequality. *Educational Research Quarterly*, *33*(1), 60-74.

Dumas, F., Huguet, P., Monteil, J., Rastoul, C., & Nezlek, J. (2005). Social comparison in the classroom: Is there a tendency to compare upward in elementary school? *Current Research in Social Psychology*, *10*(12), 166-187.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education, 54*(3), 151-162.

Gentry, M. (2006). No child left behind: neglecting excellence. *Roeper Review, 29*(1), 24-27. Retrieved from http://www.geri.education.purdue.edu/PDF %20Files/Gentry_Publications/2006._Gentry._Neglec.pdf

Haller, E., & Davis, S. (1981). Teacher perceptions, parental social status and grouping for reading instruction. *Sociology of Education*, *54*(3), 162-174.

Hulan, N. (2010). What the students will say while the teacher is away: An investigation into student-led and teacher-led discussion within guided reading groups. *Literacy Teaching and Learning, 14*(1), 41-64.

Kersten, J. (2007). Literacy and choice: Urban elementary students' perceptions of links between home, school, and community literacy practices. In V. Purcell-Gates (Ed.), *Cultural practices of literacy*. London, England: Taylor & Francis.

Krebs, S., & Roebers, C. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, *80*(3), 325-340.

Leder, G. (1987). Teacher student interaction: A case study. *Educational Studies in Mathematics, 18*(3) 255-271.

LeTendre, G., Hofer, B, & Shimizu, H. (2003). What is tracking? Cultural
        expectations in the United States, Germany, and Japan. *American*
        *Educational Research Journal*, *40*(1), 43-89.

Lopata, C., Wallace, N., & Finn, K. (2005). Comparison of academic achievement
        between Montessori and traditional education programs. *Journal of*
        *Research in Childhood Education, 20*(1), 5-13.

Loveless, T. (1998). *Making sense of the tracking and ability grouping debate*.
        Retrieved from http://www.edexcellence.net/publications/tracking.html

Lyle, S. (1999). An investigation of pupil perceptions of mixed ability grouping to
        enhance literacy in children aged 9-10. *Educational Studies*, 25(3),
        283-296.

Massachusetts Commission on Industrial and Technical Education. (1906).
        *Report of the commission on industrial and technical education.* Retrieved
        from http://archive.org/stream/cu31924021891290#page/n5/mode/2up

Master timeline: 1900. (2001) *School: The story of American public school*
        *education*. Retrieved from
        http://www.pbs.org/kcet/publicschool/roots_in_history/choicemaster3.html

McCoach, D., O'Connell, A., & Levitt, H. (2006). Ability grouping across
        kindergarten using an early childhood longitudinal study. *The Journal of*
        *Educational  Research*, *99*(6), 339-346.

Monsaas, J. A., & Engelhard, G., Jr. (1994). Teachers' attitudes toward testing
        practices. *The Journal of Psychology, 128*(4), 469-77.

National Education Association. (2010). *Research spotlight on academic ability*
        *grouping*. Retrieved from http://www.nea.org/tools/16899.htm

Neumeister, K., Adams, C., Pierce, R., Cassady, J., & Dixon, F. (2007). Fourth
        grade teachers' perceptions of giftedness: Implications for identifying and
        serving diverse gifted students. *Journal for the Education of the Gifted*,
        *30*(4), 479-499.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed). New
      Haven, Connecticut: Yale University Press.

Office of Superintendent of Public Instruction. (2010). *Personnel by major
      position and ethnicity*. Retrieved from
      http://www.k12.wa.us/DataAdmin/pubdocs/personnel/positionandethnicity0
      910.pdf

Parsons, J., Kaczala, C., & Meece, J. (1982). Socialization of achievement
      attitudes and beliefs: Classroom influences. *Child Development*, 53(2),
      322-339.

Patrick, R., Snyder, J., Schrepferman, L., & Snyder, J. (2005). The joint
      contribution of early parental warmth, communication and tracking, and
      early child conduct problems on monitoring in late childhood. *Child
      Development, 76*(5), 999-1014.

Skybo, T., & Buck, J. (2007). Stress and coping responses to proficiency testing
      in school-age children. *Pediatric Nursing, 33*(5), 413-8.

Sørenson, A., & Hallinan, M. (1986). Effects of ability grouping on growth in
      academic achievement. *American Educational Research Journal, 23*(4),
      519-542.

Spring, J. (2010). *The American school, a global context: From the Puritans to
      the Obama administration* (8th ed). New York, New York: McGraw-Hill.

Stevens, R., & Slavin, R. (1995). The cooperative elementary school: Effects on
      students' achievement, attitudes, and social relations. *American
      Educational Research Journal, 32*(2), 321-351.

Thompson, S., Meyers, J., & Oshima, T. (2011). Student mobility and its
      implications for schools' adequate yearly progress. *The Journal of Negro
      Education, 80*(1), 12-21.

U.S. Census Bureau, Population Division (2000). *Historical national population
      estimates: July 1, 1900 to July 1, 1999.* Retrieved from
      http://www.census.gov/popest/archives/1990s/popclockest.txt

Waxman, H., Huang, S., Anderson, L., & Weinstein, T. (1997). Classroom process differences in inner-city elementary schools. *Journal of Educational Research, 91*(1), 49-59.

Welner, K., & Burris, C. (2006). Alternative approaches to the politics of detracking. *Theory into Practice, 45*(1), 90-99.

Worthy, J. (2010). Only the names have changed: Ability grouping revisited. *The Urban Review, 42*(4), 271-295.

Yonezawa, S., Wells, A., & Serna, I. (2002). Choosing tracks: "Freedom of choice" in detracking schools. *American Educational Research Journal, 39*(1), 37-67.